



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850

Citation for published version:

Brohan, P, Kennedy, JJ, Harris, I, Tett, SFB & Jones, PD 2006, 'Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850', *Journal of Geophysical Research*, vol. 111, no. D12, D12106, pp. 1-21. <https://doi.org/10.1029/2005JD006548>

Digital Object Identifier (DOI):

[10.1029/2005JD006548](https://doi.org/10.1029/2005JD006548)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Journal of Geophysical Research

Publisher Rights Statement:

Published in the Journal of Geophysical Research: Atmospheres by the American Geophysical Union (2006)

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850

P. Brohan,¹ J. J. Kennedy,¹ I. Harris,² S. F. B. Tett,³ and P. D. Jones²

Received 2 August 2005; revised 19 December 2005; accepted 14 February 2006; published 24 June 2006.

[1] The historical surface temperature data set HadCRUT provides a record of surface temperature trends and variability since 1850. A new version of this data set, HadCRUT3, has been produced, benefiting from recent improvements to the sea surface temperature data set which forms its marine component, and from improvements to the station records which provide the land data. A comprehensive set of uncertainty estimates has been derived to accompany the data: Estimates of measurement and sampling error, temperature bias effects, and the effect of limited observational coverage on large-scale averages have all been made. Since the mid twentieth century the uncertainties in global and hemispheric mean temperatures are small, and the temperature increase greatly exceeds its uncertainty. In earlier periods the uncertainties are larger, but the temperature increase over the twentieth century is still significantly larger than its uncertainty.

Citation: Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res.*, *111*, D12106, doi:10.1029/2005JD006548.

1. Introduction

[2] The historical surface temperature data set HadCRUT [Jones, 1994; Jones and Moberg, 2003] has been extensively used as a source of information on surface temperature trends and variability [Houghton *et al.*, 2001]. Since the last update, which produced HadCRUT2 [Jones and Moberg, 2003], important improvements have been made in the marine component of the data set [Rayner *et al.*, 2006]. These include the use of additional observations, the development of comprehensive uncertainty estimates, and technical improvements that enable, for instance, the production of gridded fields at arbitrary resolution.

[3] This paper describes work to produce a new data set version, HadCRUT3, which will extend the advances made to the marine data to the global data set. These new developments include improvements to: the land station data, the process for blending land data with marine data to give global coverage, and the statistical process of adjusting the variance of the gridded values to allow for varying numbers of contributing observations. Results and uncertainties for the new blended, global data set, called HadCRUT3, are presented.

2. Land Surface Data

2.1. Station Data

[4] The land surface component of HadCRUT is derived from a collection of homogenized, quality-controlled, monthly averaged temperatures for 4349 stations. This collection has been expanded and improved for use in the new data set.

2.1.1. Additional Stations and Data

[5] New stations and data were added for Mali, the Democratic Republic of Congo, Switzerland [Begert *et al.*, 2005] and Austria. Data for 16 Austrian stations were completely replaced with revised values. A total of 29 Mali series were affected: 5 had partial new data, 8 had completely new data, and 16 were new stations. Five Swiss stations were updated for the period 1864–2001 [Begert *et al.*, 2005]. Thirty-three Congolese stations were affected: Thirteen were new stations, and 20 were updates to existing stations.

[6] As well as the new stations discussed above, additional monthly data have been obtained for stations in Antarctica [Turner *et al.*, 2005], while additional data for many stations have been added from the National Climatic Data Centre publication Monthly Climatic Data for the World.

2.1.2. Quality Control

[7] Much additional quality control has also been undertaken. A comparison [Simmons *et al.*, 2004] of the Climatic Research Unit (CRU) land temperature data with the ERA-40 reanalysis found a few areas where the station data were doubtful, and this was augmented by visual examination of individual station records looking for outliers. Some bad values were identified and either corrected or removed. Only a small fraction of the data needed correction, however; of the more than 3.7 million

¹Hadley Centre for Climate Prediction and Research, Met Office, Exeter, UK.

²Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, UK.

³Met Office Hadley Centre (Reading Unit), University of Reading, Reading, UK.

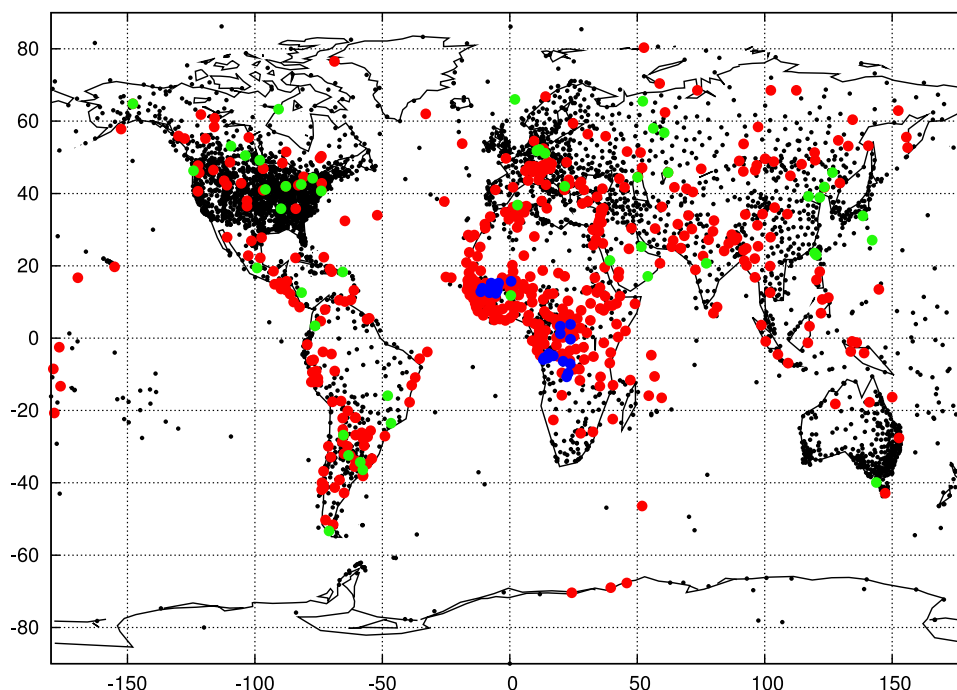


Figure 1. Land station coverage. Black circles mark all stations, green circles mark deleted stations, blue circles mark stations added, and red circles mark stations edited. Many station edits are minor changes, involving, for instance, the correction of a single outlier.

monthly station values, the ERA-40 comparison found about 10 doubtful grid boxes and the visual inspection about 270 monthly outliers.

[8] Checking the station data for identical sequences in all possible station pairs turned up 53 stations which were duplicates of others. These duplicates have arisen where the same station data are assimilated into the archive from two different sources, and the two sources give the same station but with different names and WMO identifiers. The duplicate stations were merged and duplicate temperature data were deleted.

[9] Also the station normals and standard deviations were improved. The station normals (monthly averages over the normal period 1961–1990) are generated from station data for this period where possible. Where there are insufficient station data to achieve this for the period, normals were derived from WMO values [World Meteorological Organization (WMO), 1996] or inferred from surrounding station values [Jones *et al.*, 1985]. For 617 stations, it was possible to replace the additional WMO normals (used by Jones and Moberg [2003]) with normals derived from the station data. This was made possible by relaxing the requirement to have data for 4 years in each of the three decades in 1961–1990 (the requirement now is simply to have at least 15 years of data in this period), so reducing the number of stations using the seemingly less reliable WMO normals. As well as making the normals less uncertain (see the discussion of normal error below), these improved normals mean that the gridded fields of temperature anomalies are much closer to zero over the normal period than was the case for previous versions of the data set. Figure 1 shows the locations of the stations used, and indicates those where changes have been made.

2.2. Gridding

[10] To interpolate the station data to a regular grid the methods of [Jones and Moberg, 2003] are followed. Each grid box value is the mean of all available station anomaly values, except that station outliers in excess of five standard deviations are omitted.

[11] Two changes have been made in the gridding process. The station anomalies can now be gridded to any spatial resolution, instead of being limited to a $5^\circ \times 5^\circ$ resolution; this simplifies comparison of the gridded data with General Circulation Model (GCM) results. Also previous versions of the data set did some infilling of missing grid box values using data from surrounding grid boxes [Jones *et al.*, 2001]. This is no longer done, allowing the attribution of an uncertainty to each grid box value. The resulting gridded land-only data set has been given the name CRUTEM3. The previous version of this data set, CRUTEM2, started in 1851: In CRUTEM3 the start date has been extended back to 1850 to match the marine data (section 3). Figure 2 shows a gridded field for an example month, at the standard $5^\circ \times 5^\circ$ degree resolution.

[12] For comparison with GCM results, or for regional studies of areas where observations are plentiful, it can be useful to perform the gridding at higher resolution. Figure 3 shows a gridded field for the same example month, at the resolution of the HadGEM1 model [Johns *et al.*, 2004], but only for North America.

2.3. Uncertainties

[13] To use the data for quantitative, statistical analysis, for instance, a detailed comparison with GCM results, the uncertainties of the gridded anomalies are a useful additional field. A definitive assessment of uncertainties is

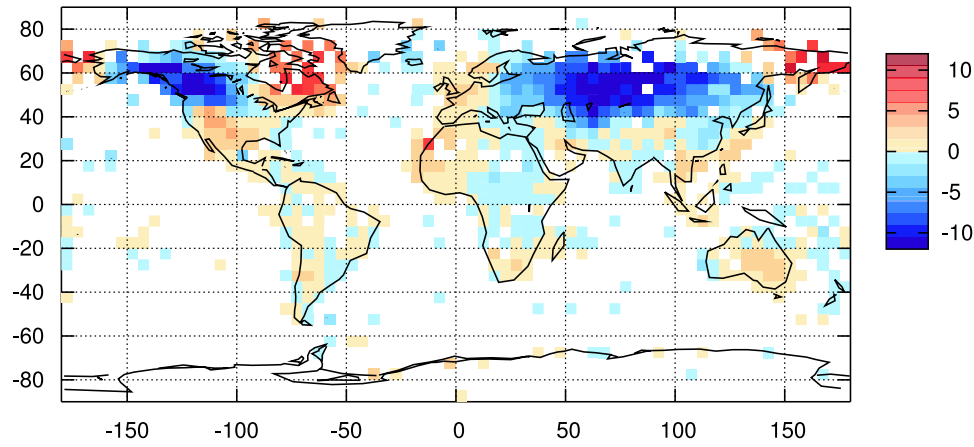


Figure 2. CRUTEM3 anomalies ($^{\circ}\text{C}$) for January 1969 (global, $5^{\circ} \times 5^{\circ}$).

impossible, because it is always possible that some unknown error has contaminated the data, and no quantitative allowance can be made for such unknowns. There are, however, several known limitations in the data, and estimates of the likely effects of these limitations can be made (Defense secretary Rumsfeld press conference, June 6, Back to disarmament documentation, June 2002, London, The Acronym Institute (available at www.acronym.org.uk/docs/0206/doc04.htm)). This means that uncertainty estimates need to be accompanied by an error model: a precise description of what uncertainties are being estimated.

[14] Uncertainties in the land data can be divided into three groups: (1) station error, the uncertainty of individual station anomalies; (2) sampling error, the uncertainty in a grid box mean caused by estimating the mean from a small number of point values; and (3) bias error, the uncertainty in large-scale temperatures caused by systematic changes in measurement methods.

2.3.1. Station Errors

[15] The uncertainties in the reported station monthly mean temperatures can be further sub divided. Suppose

$$T_{\text{actual}} = T_{\text{ob}} + \epsilon_{\text{ob}} + C_H + \epsilon_H + \epsilon_{RC}, \quad (1)$$

where T_{actual} is the actual station mean monthly temperature, T_{ob} is the reported temperature, ϵ_{ob} is the measurement error, C_H is any homogenization adjustment that may have

been applied to the reported temperature and ϵ_H is the uncertainty in that adjustment, and ϵ_{RC} is the uncertainty due to inaccurate calculation or miss reporting of the station mean temperature.

[16] The values being gridded are anomalies, calculated by subtracting the station normal from the observed temperature, so errors in the station normals must also be considered.

$$A_{\text{actual}} = T_{\text{ob}} - T_N + \epsilon_N + \epsilon_{\text{ob}} + C_H + \epsilon_H + \epsilon_{RC}, \quad (2)$$

where A_{actual} is the actual temperature anomaly, T_N is the estimated station normal, and ϵ_N is the error in T_N .

[17] The basic station data include normals and may have had homogenization adjustments applied, so they provide $T_{\text{ob}} + C_H$ and T_N ; also needed are estimates for ϵ_{ob} , ϵ_H , ϵ_N , and ϵ_{RC} .

2.3.1.1. Measurement Error (ϵ_{ob})

[18] The random error in a single thermometer reading is about 0.2°C (1σ) [Folland *et al.*, 2001]; the monthly average will be based on at least two readings a day throughout the month, giving 60 or more values contributing to the mean. So the error in the monthly average will be at most $0.2/\sqrt{60} = 0.03^{\circ}\text{C}$ and this will be uncorrelated with the value for any other station or the value for any other month.

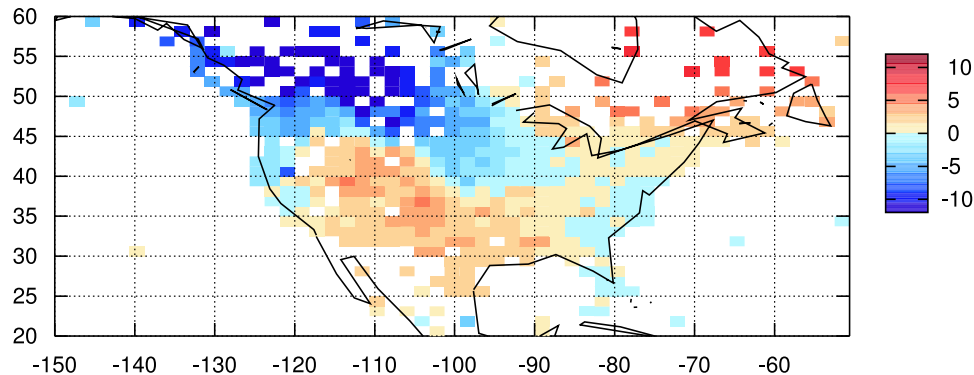


Figure 3. CRUTEM3 anomalies ($^{\circ}\text{C}$) for January 1969 (North America, HadGEM1 model grid ($1.875^{\circ} \times 1.25^{\circ}$)).

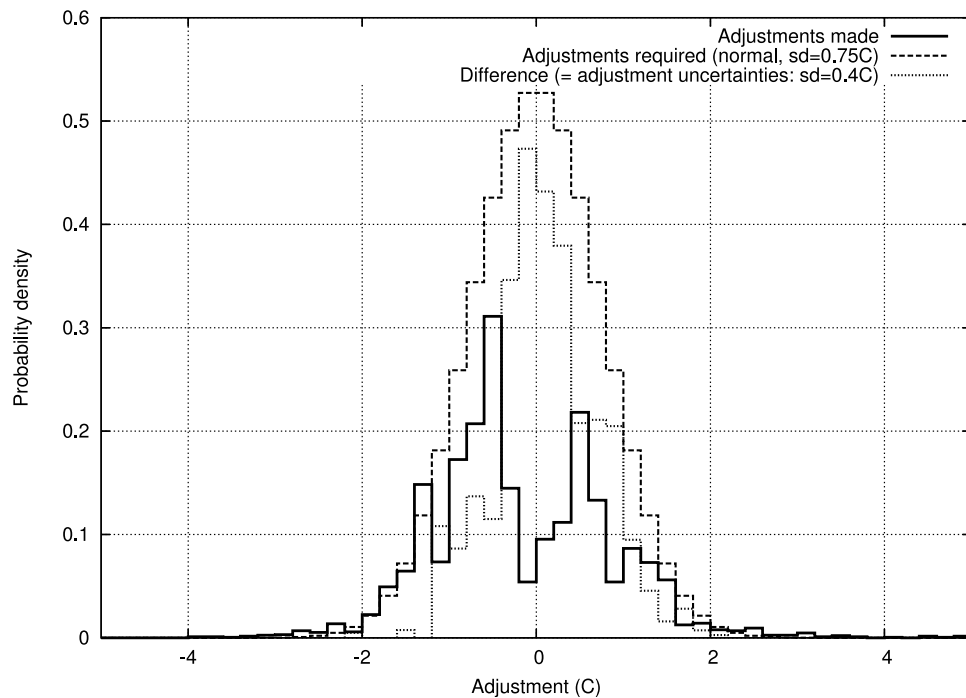


Figure 4. Distribution of station homogeneity adjustments ($^{\circ}\text{C}$). The solid line is the distribution of the adjustments known to have been made (763 adjustments from *Jones et al.* [1985, 1986] and *Vincent and Gullet* [1999]), the dashed line is a hypothesized distribution of the adjustments required, and the dotted line is the difference and so the distribution of homogeneity adjustment error.

[19] There will be a difference between the true mean monthly temperature (i.e., from 1 min averages) and the average calculated by each station from measurements made less often; but this difference will also be present in the station normal and will cancel in the anomaly. So this does not contribute to the measurement error. If a station changes the way mean monthly temperature is calculated it will produce an inhomogeneity in the station temperature series, and uncertainties due to such changes will form part of the homogenization adjustment error.

2.3.1.2. Homogenization Adjustment Error (ϵ_H)

[20] Inhomogeneities are introduced into the station temperature series by such things as changes in the station site, changes in measurement time, or changes in instrumentation. The station data that are used to make HadCRUT have been adjusted to remove these inhomogeneities, but such adjustments are not exact; there are uncertainties associated with them.

[21] For some stations both the adjusted and unadjusted time series are archived at CRU and so the adjustments that have been made are known [*Jones et al.*, 1985, 1986; *Vincent and Gullet*, 1999], but for most stations only a single series is archived, so any adjustments that might have been made (e.g., by National Met. services or individual scientists) are unknown.

[22] Making a histogram of the adjustments applied (where these are known) gives the solid line in Figure 4. Inhomogeneities will come in all sizes, but large inhomogeneities are more likely to be found and adjusted than small ones. So the distribution of adjustments is bimodal, and can be interpreted as a bell-shaped distribution with most of the central, small, values missing.

[23] Hypothesizing that the distribution of adjustments required is Gaussian, with a standard deviation of 0.75°C gives the dashed line in Figure 4 which matches the number of adjustments made where the adjustments are large, but suggests a large number of missing small adjustments. The homogenization uncertainty is then given by this missing component (dotted line in Figure 4), which has a standard deviation of 0.4°C . This uncertainty applies to both adjusted and unadjusted data, the former have an uncertainty on the adjustments made, the latter may require undetected adjustments.

[24] The distribution of known adjustments is not symmetric; adjustments are more likely to be negative than positive. The most common reason for a station needing adjustment is a site move in the 1940–1960 period. The earlier site tends to have been warmer than the later one, as the move is often to an out of town airport. So the adjustments are mainly negative, because the earlier record (in the town/city) needs to be reduced [*Jones et al.*, 1985, 1986]. Although a real effect, this asymmetry is small compared with the typical adjustment, and is difficult to quantify; so the homogenization adjustment uncertainties are treated as being symmetric about zero.

[25] The homogenization adjustment applied to a station is usually constant over long periods: The mean time over which an adjustment is applied is nearly 40 years [*Jones et al.*, 1985, 1986; *Vincent and Gullet*, 1999]. The error in each adjustment will therefore be constant over the same period. This means that the adjustment uncertainty is highly correlated in time: The adjustment uncertainty on a station value will be the same for a decadal average as for an individual monthly value.

[26] So the homogenization adjustment uncertainty for any station is a random value taken from a normal distribution with a standard deviation of 0.4°C . Each station uncertainty is constant in time, but uncertainties for different stations are not correlated with one another (correlated inhomogeneities are treated as biases, see below). As an inhomogeneity is a change from the conditions over the climatology period (1961–1990), station anomalies will have no inhomogeneities during that period unless there is a change sometime during those 30 years. Consequently these adjustment uncertainty estimates are pessimistic for that period.

[27] Figure 4 also demonstrates the value of making homogenization adjustments. The dashed line is an estimate of the uncertainties in the unadjusted data, and the dotted line an estimate of the uncertainties remaining after adjustment. The adjustments made have reduced the uncertainties considerably.

2.3.1.3. Normal Error (ϵ_N)

[28] For most stations, the station normal is calculated from the monthly temperatures for that station over the normal period (1961–1990). So the uncertainty in the normal consists of measurement and sampling error for that data. The measurement error will be a small fraction of the monthly measurement error and can be neglected, so only the sampling error is important.

[29] The station temperature in each month during the normal period can be considered as the sum of two components: a constant station normal value (C) and a random weather value (w , with standard deviation σ_i). If data for a station are available for N of the 30 possible months during the period from which the normals are taken, and the w s are uncorrelated; then for stations where C is estimated as the mean of the available monthly data, the uncertainty on C is σ_i/\sqrt{N} . Testing this model by selecting stations where complete data are available for the climatology period and looking at the effect on the normals of using only a subset of the data confirmed that the autocorrelation is small and the model is appropriate.

[30] The station normals used fall into three groups [Jones and Moberg, 2003]. The first group are those where data are available for all months in 1961–1990; these normals are given an uncertainty of $\sigma_i/\sqrt{30}$. The second group are those where data are available for at least 15 years in 1961–1990 (enough data to estimate a normal); these normals are given an uncertainty of σ_i/\sqrt{N} where N is the number of years for which there is data. The third group are those where too few data are available in 1961–1990 to estimate a normal. For some of these stations WMO normals have been used [WMO, 1996] and experience has shown that these normals are likely to have problems [Jones and Moberg, 2003]. The process of data improvement discussed in section 2.1.2 also allowed the generation of new normals for 617 such stations. Comparison of the old and new normals for these stations suggested that the uncertainty in the WMO normals was about $0.3\sigma_i$.

2.3.1.4. Calculation and Reporting Error (ϵ_{RC})

[31] The station data used in this analysis may have been extensively processed before being added to the CRU archive. The monthly mean temperature values will have been calculated by averaging 60 or more subdaily measurements. Where this calculation is done manually it can

introduce an error. The transmission of the station data to the CRU archive requires at least one cycle of encoding, transcribing and decoding the data, and this process may also introduce an error.

[32] Where such errors are persistent they will introduce an inhomogeneity into the data for a station, and so are included in the homogenization adjustment error ϵ_H . So the calculation and reporting error (ϵ_{RC}) is composed of only the random and uncorrelated cases.

[33] Calculation and reporting errors can be large (changing the sign of a number and scaling it by a factor of 10 are both typical transcription errors; as are reporting errors of 10°C (e.g., putting 29.1 for 19.1)) but almost all such errors will be found during quality control of the data. Those errors that remain after quality control will be small, and because they are also uncorrelated both in time and in space their effect on any large-scale average will be negligible. For these reasons ϵ_{RC} is not considered further.

2.3.1.5. Combining Station Error Components

[34] For each station, the observational, homogeneity adjustment, and normal uncertainties are independent; so estimates of each can be combined in quadrature to give an estimate of the total uncertainty for each station. The grid box anomaly is the mean of the n station anomalies in that grid box, so the grid box station uncertainty is the root mean square of the station errors, multiplied by $1/\sqrt{n}$. The spatial patterns visible in the station error field (Figure 5) are dominated by the distribution of the mean station standard deviation. This is larger in the high latitudes and in the winter, and smaller in the tropics and in the summer; so for the month shown (January) the station error is largest for the northern high latitudes. A secondary effect is a reduction in areas with a large number of observations. In North America, Europe, and southeastern Australia, observations are plentiful and so the station error is reduced.

2.3.2. Sampling Error

[35] Even if the station temperature anomalies had no error, the mean of the station anomalies in a grid box would not necessarily be equal to the true spatial average temperature anomaly for that grid box. This difference is the sampling error; and it will depend on the number of stations in the grid box, on the positions of those stations, and on the actual variability of the climate in the grid box. A method for calculating sampling error is described by Jones *et al.* [1997], who recommend the equation

$$SE^2 = \frac{\overline{\sigma_i^2} \bar{r} (1 - \bar{r})}{1 + (n - 1) \bar{r}}, \quad (3)$$

where $\overline{\sigma_i^2}$ is the mean station standard deviation, n is the number of stations, and \bar{r} is the average intersite correlation (itself estimated from the data according to the methods of Jones *et al.* [1997]). The method of Jones *et al.* [1997] has been used in this analysis.

[36] The spatial distribution of sampling error (see Figure 6), like the station error, is dominated by the station standard deviations and the number of observations. The distribution is very similar to that for the station error.

2.3.3. Bias Error

[37] Bias correction uncertainties are estimated following Folland *et al.* [2001], who considered two biases in the land

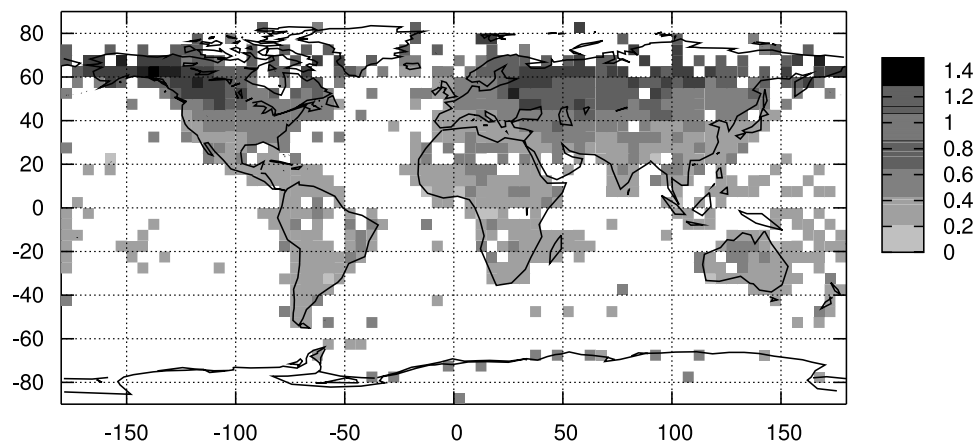


Figure 5. CRUTEM3 station errors ($^{\circ}\text{C}$) for January 1969.

data: urbanization effects [Jones *et al.*, 1990] and thermometer exposure changes [Parker, 1994].

2.3.3.1. Urbanization Effects

[38] The previous analysis of urbanization effects in the HadCRUT data set [Folland *et al.*, 2001] recommended a 1σ uncertainty which increased from 0 in 1900 to 0.05°C in 1990 (linearly extrapolated after 1990) [Jones *et al.*, 1990]. Since then, research has been published suggesting both that the urbanization effect is too small to detect [Parker, 2004; Peterson, 2004], and that the effect is as large as $\approx 0.3^{\circ}\text{C}/\text{century}$ [Kalnay and Cai, 2003; Zhou *et al.*, 2004].

[39] The studies finding a large urbanization effect [Kalnay and Cai, 2003; Zhou *et al.*, 2004] are based on comparison of observations with reanalyses, and assume that any difference is entirely due to biases in the observations. A comparison of HadCRUT data with the ERA-40 reanalysis [Simmons *et al.*, 2004] demonstrated that there were sizable biases in the reanalysis, so this assumption cannot be made, and the most reliable way to investigate possible urbanization biases is to compare rural and urban station series.

[40] A recent study of rural/urban station comparisons [Peterson and Owen, 2005] supported the previously used recommendation [Jones *et al.*, 1990], and also demonstrated that assessments of urbanization were very dependent on the choice of metadata used to make the rural/urban classifica-

tion. To make an urbanization assessment for all the stations used in the HadCRUT data set would require suitable metadata for each station for the whole period since 1850. No such complete metadata are available, so in this analysis the same value for urbanization uncertainty is used as in the previous analysis [Folland *et al.*, 2001]; that is, a 1σ value of $0.0055^{\circ}\text{C}/\text{decade}$, starting in 1900. Recent research suggests that this value is reasonable, or possibly a little conservative [Parker, 2004; Peterson, 2004; Peterson and Owen, 2005]. The same value is used over the whole land surface, and it is one-sided: Recent temperatures may be too high because of urbanization, but they will not be too low.

2.3.3.2. Thermometer Exposure Changes

[41] Over the period since 1850 there have been changes in the design and siting of thermometer enclosures; many early shelters can differ substantially from the modern Stevenson-type screen. It is sometimes possible to determine the time of change by the homogeneity assessments discussed in section 2.3.1.2, but this is only possible if changes at neighboring stations are implemented at different times. The bias errors in this section therefore allow for the possible simultaneous replacement across entire countries with Stevenson-type shelters. The possible effect of such changes was investigated by Parker [1994], who concluded that there was a possible difference between 1900 and the present day of about 0.2°C because of such exposure

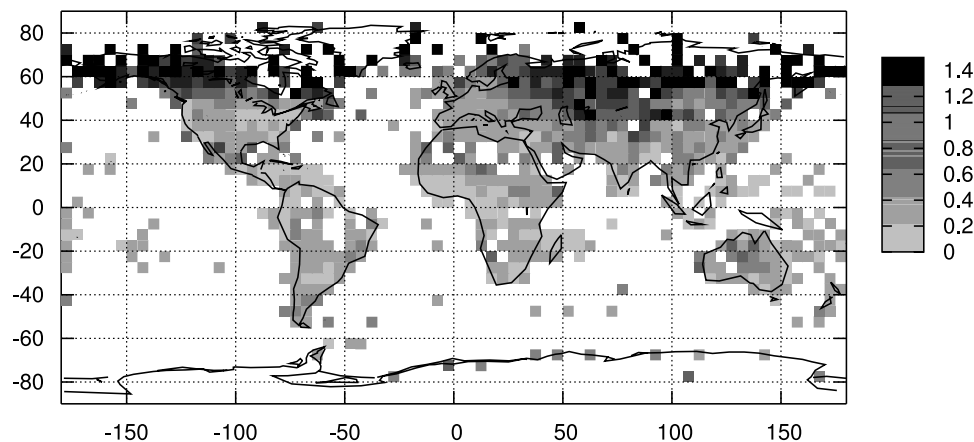


Figure 6. CRUTEM3 sampling errors ($^{\circ}\text{C}$) for January 1969.

changes. This was later expanded into an error model by *Folland et al.* [2001]: In the tropics (20S–20N) the 1σ uncertainty range is 0.2°C before 1930, and then decreases linearly to zero in 1950. Outside the tropics the 1σ uncertainty range is 0.1°C before 1900 and then decreases linearly to zero by 1930. This uncertainty model is used here.

[42] It is likely that further changes in thermometer exposure have been taking place in recent years, as Stevenson-type screens are replaced with aspirated shelters. These changes are, however, too recent to allow a quantitative assessment of their effects and they are not included in the CRUTEM3 error analysis.

2.3.4. Combining the Uncertainties

[43] The total uncertainty value for any grid box can be obtained by adding the station error, sampling error, and bias error estimates for that grid box in quadrature. This gives the total uncertainty for each grid box for each month.

[44] In practice, however, this combined uncertainty is less useful than the individual components. Most uses of the data set require not just an individual monthly grid box value but some spatial or temporal average of many of them. When combining uncertainties onto these larger scales it is necessary to allow for correlations between the grid box uncertainties, and the three error components have different spatial and temporal correlation structures.

[45] The sampling errors have little spatial or temporal correlation. The station errors have little spatial correlation, but because the two main components (homogeneity adjustment and normal uncertainties) stay the same for each station for many consecutive months they have almost complete temporal autocorrelation. The bias errors are the same for each grid box and each month, they have complete temporal and spatial correlations.

[46] The errors shown in Figures 5 and 6 are for $5^{\circ} \times 5^{\circ}$ grid boxes. Changing the gridding resolution will change the uncertainties. Larger grid boxes will have a larger sampling error if they contain the same number of observations, but typically increasing the grid box size will mean that each contains more stations and the box-averaged uncertainties will be reduced. Similarly, reducing the grid box size would reduce the sampling error, except that smaller grid boxes will often contain fewer stations, which will increase the errors.

[47] The combined effect of grid box sampling errors will be small for any continental-scale or hemispheric-scale average (though the lack of global coverage introduces an additional source of sampling error, this is discussed in section 6.1). Combined station errors will be small for large-scale spatial averages, but remain important for averages over long periods of the same small grid box. Bias errors are equally large on any space or timescale.

3. Marine Data

[48] The marine data used are from the sea surface temperature data set HadSST2 [*Rayner et al.*, 2006]. This is a gridded data set made from in situ ship and buoy observations from the new International Comprehensive Ocean-Atmosphere data set [*Diaz et al.*, 2002; *Manabe*, 2003; *Woodruff et al.*, 2003]. This data set provides the same information for the oceans as described above for the

land. For each grid box, mean temperature anomalies, measurement and sampling error estimates, and bias error estimates are available. The data sets can be produced on a grid of any desired resolution.

[49] Previous versions of HadCRUT use the SST data set MOHSST6 [*Parker et al.*, 1995]. The new HadSST2 data set is an improvement on MOHSST6 for many reasons: It is based on an enlarged and improved set of ship and buoy observations, it includes a new climatology, and the bias corrections needed for data before 1941 have been revisited. Also HadSST2 starts in January 1850 (as does HadCRUT3), MOHSST6 and HadCRUT2 started in January 1856. Full details of all the improvements are given by *Rayner et al.* [2006].

[50] Blending a sea surface temperature (SST) data set with land air temperature makes an implicit assumption that SST anomalies are a good surrogate for marine air temperature anomalies. It has been shown, for example, by *Parker et al.* [1994], that this is the case, and that marine SST measurements provide more useful data and smaller sampling errors than marine air temperature measurements would. So blending SST anomalies with land air temperature anomalies is a sensible choice.

[51] Like the land data, the marine data set has known errors: Estimates have been made of the measurement and sampling error, and the uncertainty in the bias corrections. The marine data are point measurements from moving ships, moored buoys, and drifting buoys, so the anomalies for any one grid box come in general from a different set of sources each month. This means that marine data have no equivalent of station errors or homogenization adjustments. The marine equivalent of the station errors form part of the measurement and sampling error, and adjustments for inhomogeneities are done by large-scale bias corrections.

[52] The measurement and sampling error estimates are based, like the land sampling error (section 2.3.2), on the number of observations in a grid box, on the variability of a single observation, and on the correlation between observations. The latter two parameters are estimated from the gridded data for each grid box. Details are given by *Rayner et al.* [2006].

[53] Only one bias correction is applied: Over the period 1850–1940, the predominant SST measurement process changed from taking samples in wooden buckets, to taking samples in canvas buckets, to using engine room cooling water inlet temperatures [*Folland and Parker*, 1995]. A bias correction is applied to remove the effect of these changes on the SSTs. This correction depends on estimates of the mix of measuring methods in use at any one time, and of parameters such as the speed of the ships making the measurements. An uncertainty has been estimated for the correction; again, details are given by *Rayner et al.* [2006].

[54] As with the land data, the uncertainty estimates cannot be definitive: Where there are known sources of uncertainty, estimates of the size of those uncertainties have been made. There may be additional sources of uncertainty as yet unquantified (see section 6.3).

4. Blending Land and Marine Data

[55] To make a data set with global coverage the land and marine data must be combined. For land-only grid boxes the

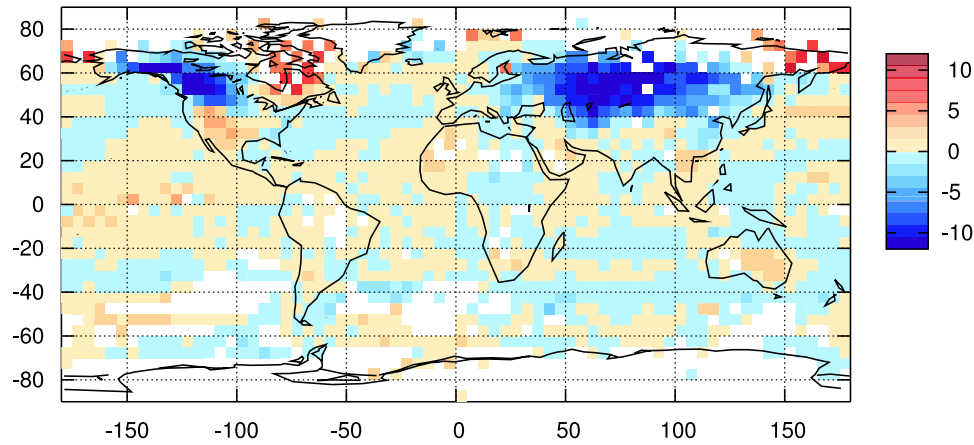


Figure 7. HadCRUT3 anomalies ($^{\circ}\text{C}$) for January 1969.

land value is taken, and for sea-only grid boxes the marine value; but for coastal and island grid boxes the land and marine data must be blended into a combined average.

[56] Previous versions of HadCRUT [Jones, 1994; Jones and Moberg, 2003] blended land and sea data in coastal and island grid boxes by weighting the land and sea values by the area fraction of land and sea respectively, with a constraint that the land fraction cannot be greater than 75% or less than 25%, to prevent either data source being swamped by the other. The aim of weighting by area was to place more weight on the more reliable data source where possible. The constraints are necessary because there are some grid boxes which are almost all sea but contain one reliable land station on a small island; and some grid boxes which are almost all land but also include a small sea area which has many marine observations. Unconstrained weighting by area would essentially discard one of the measurements, which is undesirable.

[57] The new developments described in this paper provide measurement and sampling uncertainty estimates for each grid box in both the land and marine data sets. This means that the land and marine data can be blended in the way that minimizes the uncertainty of the blended mean. That is, by scaling according to their uncertainties,

so that the more reliable value has a higher weighting than the less reliable.

$$T_{\text{blended}} = \frac{\epsilon_{\text{sea}}^2 T_{\text{land}} + \epsilon_{\text{land}}^2 T_{\text{sea}}}{\epsilon_{\text{land}}^2 + \epsilon_{\text{sea}}^2}, \quad (4)$$

where T_{blended} is the blended average temperature anomaly, T_{land} and T_{sea} are the land and marine anomalies, ϵ_{land} is the measurement and sampling error of the land data, and ϵ_{sea} is the measurement and sampling error of the marine data.

[58] The resulting blended data set for a sample month (Figure 7) shows the coherency between the land and sea data: Large-scale regions of positive or negative temperature anomalies that cross land-sea boundaries show up clearly. The land data weighting for all coastal and island grid boxes with both land and sea data for the same month (Figure 8) shows that weighting by uncertainties generally weights the marine data more highly where the marine data are expected to be good (North Atlantic and North Pacific coasts where there are many marine observations); and similarly weights the land data more highly where it is the more reliable (in the Southern Hemisphere, notably in Indonesia and the South Pacific where marine observations

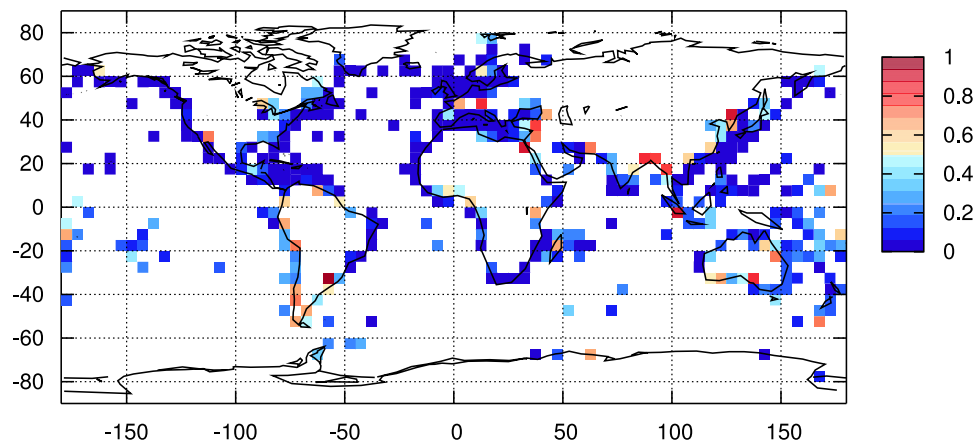


Figure 8. Land data blending weight for January 1969. (Greater emphasis on the land would give numbers closer to one).

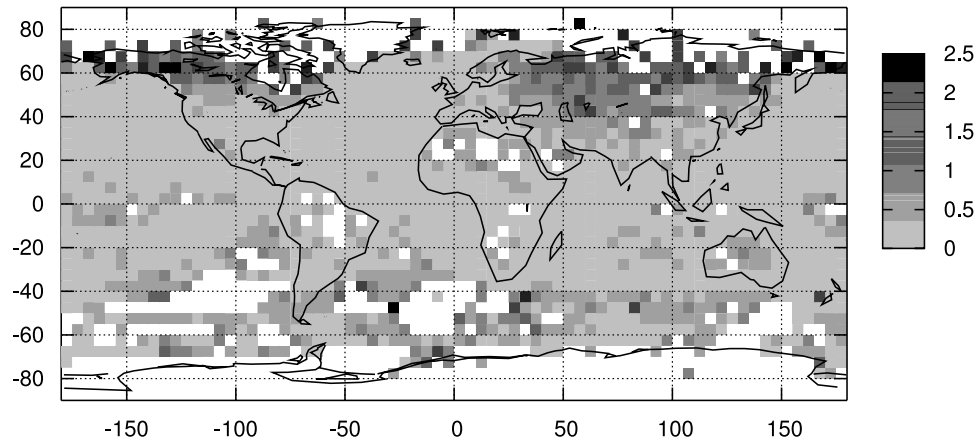


Figure 9. HadCRUT3 measurement and sampling error ($^{\circ}\text{C}$) for January 1969.

are sparse). Note that the weighting is continually varying with time as the data availability changes.

[59] As the land and marine errors are independent, this choice of weighting gives the lowest measurement and sampling error for the blended mean, giving an error in the blended mean of

$$\epsilon_{\text{blended}} = \sqrt{\frac{\epsilon_{\text{sea}}^2 \epsilon_{\text{land}}^2}{\epsilon_{\text{land}}^2 + \epsilon_{\text{sea}}^2}}. \quad (5)$$

The measurement and sampling error for the blended mean (Figure 9) is the combined station and sampling error over land (Figures 5 and 6). Over the oceans the error distribution is dominated by variations in the number of observations: Where marine observations are plentiful (North Atlantic, North Pacific and the shipping lanes) the measurement and sampling error is very small; in poorly observed areas like the Southern Ocean, the error is much larger. The errors for marine grid boxes are much smaller than those for land grid boxes because SST is less variable in both space and time than land air temperature. This difference is discussed in more detail in section 6.2. The smaller SST errors mean that the blended temperatures for coastal and island grid boxes are dominated by the SST temperatures. This is reasonable if it is assumed that, in any grid box, the land temperature and SST values for that box are each estimates of the same blended temperature. In reality this may not be true (see section 6.4) and an area-weighted average might in some cases give a more physically consistent average temperature. However, the choice of blending weight makes very little difference to large-scale averages, so the extra complexity of a blending algorithm which accounts for possible land-sea temperature anomaly differences is not justified.

5. Variance Adjustment

[60] Assigning a grid box anomaly simply as the mean of the observational anomalies in that grid box produces a good estimate of the actual temperature anomaly. However, it has the disadvantage that the variance of the grid box average is not constant in time or space; grid boxes containing many observations will have a low variance, and those with few observations a larger one. For some appli-

cations this fluctuation in variance is undesirable. Heterogeneities in the variance affect estimates of the covariance matrices which are used in EOF techniques such as Optimal Averaging. They also affect analyses of extreme monthly temperatures and of changes in temperature variability through time.

[61] For these reasons, previous versions of HadCRUT have included variance adjustments [Jones *et al.*, 2001]: alternative versions of the gridded data sets with the grid box anomalies adjusted to remove the effects of changing numbers of observations. In producing a variance adjusted version of HadCRUT3 two refinements have been made: The error estimates for the gridded data have been used to devise a simpler adjustment method applicable to both land and marine data, and the adjustment process has been tested on synthetic data to ensure that it does not introduce biases into the data. Details of the adjustment method and the tests applied are given in Appendix A. Variance adjusted versions have been produced for HadCRUT3 and the marine and land data sets from which it is formed; the adjusted data sets are named HadCRUT3v, CRUTEM3v and HadSST2v. One advantage of the new adjustment method is that it can be applied to the entire data set, so the variance adjusted data sets now also start in 1850. The previous version of the variance adjusted data set, HadCRUT2v, started in 1870.

[62] Variance adjustment is successful at the individual grid box scale: Comparison with synthetic data shows that the inflation of the grid box variance caused by the limited number of observations can be removed without introducing biases into the grid box series. At larger space scales, however, variance adjustment does introduce a small bias into the data. Whether variance adjusted or unadjusted data should be used in an analysis depends on what is to be calculated. If it is necessary that grid box anomalies have a spatially and temporally consistent variance, then variance adjusted data should be used. Otherwise, better results may be obtained using unadjusted data. In particular, global and regional time series should be calculated using unadjusted data.

6. Analyses of the Gridded Data Set

[63] From the $5^{\circ} \times 5^{\circ}$ gridded data set and its comprehensive set of uncertainty estimates it is possible to calculate a large variety of climatologically interesting summary

statistics and their uncertainty ranges. Of this variety, global and regional temperature time series probably have the widest appeal, so some illustrative examples of these are presented here.

6.1. Hemispheric and Global Time Series

[64] If the gridded data had complete coverage of the globe or the region to be averaged, then making a time series would be a simple process of averaging the gridded data and making allowances for the relative sizes of the grid boxes and the known uncertainties in the data. However, global coverage is not complete even in the years with the most observations, and it is very incomplete early in the record. In general, global and regional area averages will have an additional source of uncertainty caused by missing data.

[65] To estimate the uncertainty of a large-scale average owing to missing data the effect of subsampling on a known, complete data set is used. The NCEP/NCAR reanalysis data set [Kalnay *et al.*, 1996] provides complete monthly gridded surface air temperature values for more than 50 years. To estimate the missing data uncertainty of the HadCRUT3 mean for a particular month, the reanalysis data for that calendar month in each of the 50+ years is subsampled to have the same coverage as HadCRUT3, and the difference between the complete average and the subsampled average anomaly is calculated in each of the 50+ cases. The 2.5% and 97.5% values forming the error range of the HadCRUT3 mean for that month in the record are then estimated from the standard deviation of the 50+ differences, assuming that the differences are normally distributed. This procedure has the advantage that it works for any region, so hemispheric and regional time series and their uncertainties can be calculated as easily as global series. Unlike sophisticated optimal methods such as that used by Folland *et al.* [2001], this process makes no attempt to minimize coverage uncertainties by using estimates of data covariances. This means that the precision of large-scale averages is less than that which could be achieved with a more sophisticated method. However, the simple method has the advantage that the estimated uncertainty on the large-scale average due to limited coverage is independent of all the other sources of uncertainty. So it remains straightforward to calculate both the total uncertainty on any large-scale average and all of its components (Figure 10).

[66] This approach can also be used to give coverage uncertainties on longer timescales. Annual coverage uncertainties can be made by converting both the HadCRUT3 data and the reanalysis data to annual averages and then subsampling the annual reanalysis data with the coverage of the annual HadCRUT3 data. Similarly, estimates can be made of uncertainties of coverage uncertainties for smoothed annual or decadal averages.

[67] The grid box sampling and measurement errors are greatly reduced when the gridded data are averaged into large-scale means, so the only other important uncertainty component of global and regional time series is that owing to the biases in the data. This is dealt with by making data sets with allowances for bias uncertainties incorporated. Generating averages from data sets with bias allowances set at the 2.5% and 97.5% levels provides a 95% error range from bias uncertainties in the resulting averages.

6.1.1. Global Averages

[68] The global temperature is calculated as the mean of the Northern and Southern Hemisphere series (to stop the better sampled Northern Hemisphere from dominating the average). Figure 10 shows the global temperature anomaly time series calculated from HadCRUT3 with these error components. The monthly averages are dominated by short-term fluctuations in the anomalies; combining the data into annual averages produces a clearer picture, and smoothing the annual averages with a 21-term binomial filter highlights the low-frequency components and shows the importance of the bias uncertainties.

[69] The bias uncertainties are zero over the normal period by definition. The dominant bias uncertainties are those due to bucket correction [Rayner *et al.*, 2006] and thermometer exposure changes [Parker, 1994] both of which are large before the 1940s.

[70] A notable feature of the global time series is that the uncertainties are not always larger for earlier periods than later periods. The uncertainties are smaller in the 1850s than in the 1920s, at least for the smoothed series, despite the much larger number of observations in the 1920s. The station, sampling and measurement, and coverage errors (red and green bands in Figure 10) depend on the number and distribution of the observations, and these components of the error decrease steadily with time as the number of observations increases. These components also decrease with averaging to larger space and timescales, so they are smaller in the annual than the monthly series, and smaller again in the smoothed annual series. The bias uncertainties, however, do not reduce with spatial or temporal averaging, and they are largest in the early twentieth century; so the smoothed annual series, where the uncertainty is dominated by the bias uncertainties, also has its largest uncertainty in this period.

[71] The bias uncertainties are largest in the early twentieth century for two reasons: First, the bias uncertainties in the marine data are largest then because the uninsulated canvas buckets used in that period produced larger temperature biases than the wooden buckets used earlier (see Rayner *et al.* [2006] for details). Also because the land temperature bias uncertainties (present before 1950) are larger in the tropics than the extratropics, so for these simple global averages, the bias uncertainty depends on the ratio of station coverage in the tropics to that in the extratropics, and this ratio is smaller in the 1850s than in the 1920s.

6.1.2. Hemispheric Averages

[72] Comparing the smoothed mean temperature time series for the Northern Hemisphere and Southern Hemisphere (Figure 11) shows the difference in uncertainties between the two hemispheres. The difference in the uncertainty ranges for the two series stems from the very different land/sea ratio of the two hemispheres. The Northern Hemisphere has more land, and so a larger station, sampling and measurement error (Figure 9 and section 6.2), but it has more observations and so a smaller coverage uncertainty. The bias uncertainties are also larger in the Northern Hemisphere both because it has more land (especially in the tropics where the land biases are large), and because the SST bias uncertainties are largest in the Northern Hemisphere western boundary current regions where the SST can

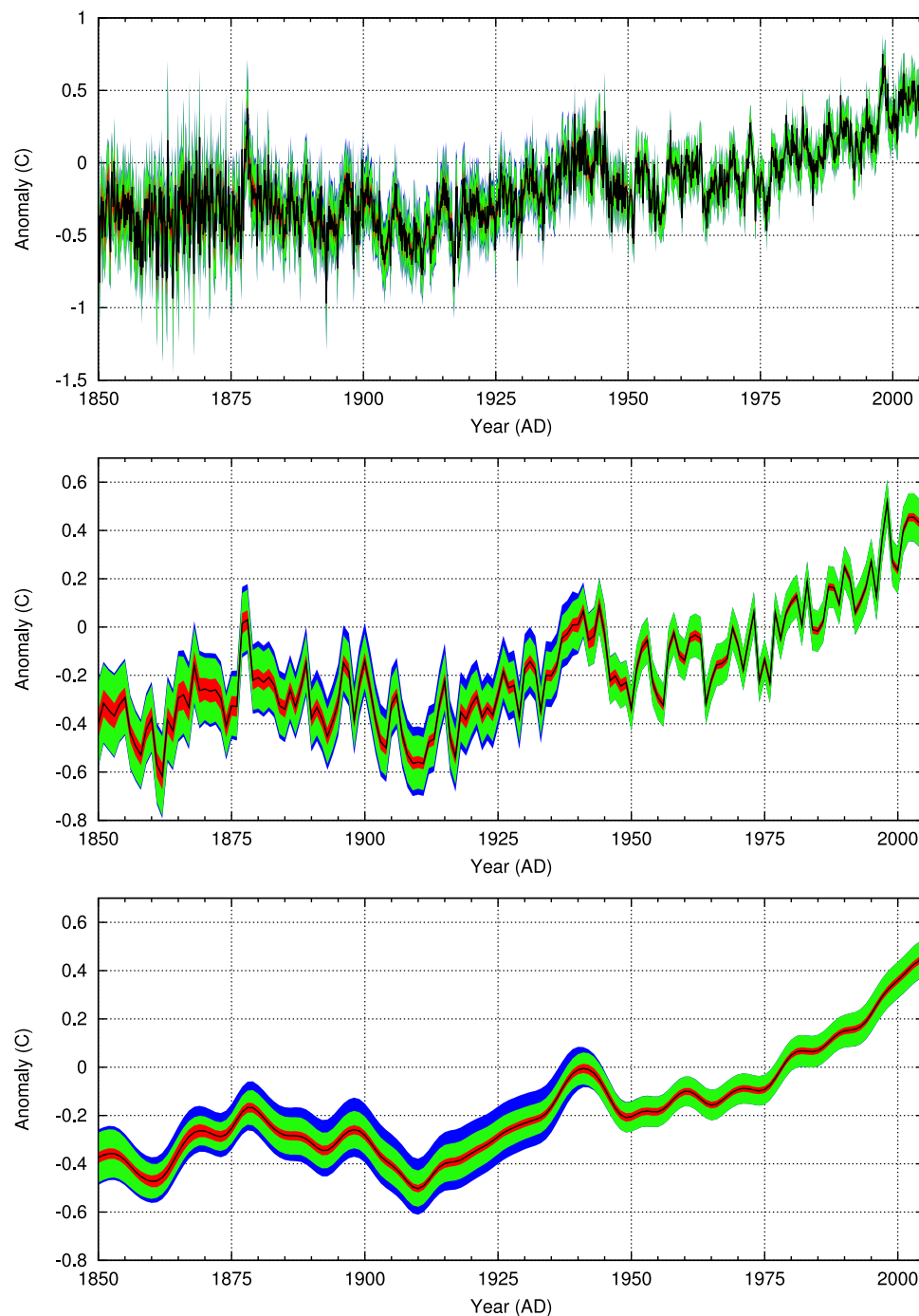


Figure 10. HadCRUT3 global temperature anomaly time series ($^{\circ}\text{C}$) at (top) monthly, (center) annual, and (bottom) smoothed annual resolutions. The black line is the best estimate value; the red band gives the 95% uncertainty range caused by station, sampling, and measurement errors; the green band adds the 95% error range due to limited coverage; and the blue band adds the 95% error range due to bias errors.

be very different from the air temperature [Rayner *et al.*, 2006].

[73] The difference between the two hemisphere series has a smaller uncertainty than either hemispheric value over much of the period shown, because the bias errors, though unknown, will be much the same in the two hemispheres and so mostly cancel in the difference. So the previously observed increase in the interhemispheric difference in the

mid twentieth century [see, e.g., Folland *et al.*, 1986; Kerr, 2005] is shown to be significantly outside the uncertainties.

6.2. Differences Between Land and Marine Data

[74] Comparison of global average time series for land-only and marine-only data (Figure 12) demonstrates both a marked agreement in the temperature trends, and a large difference in the uncertainties.

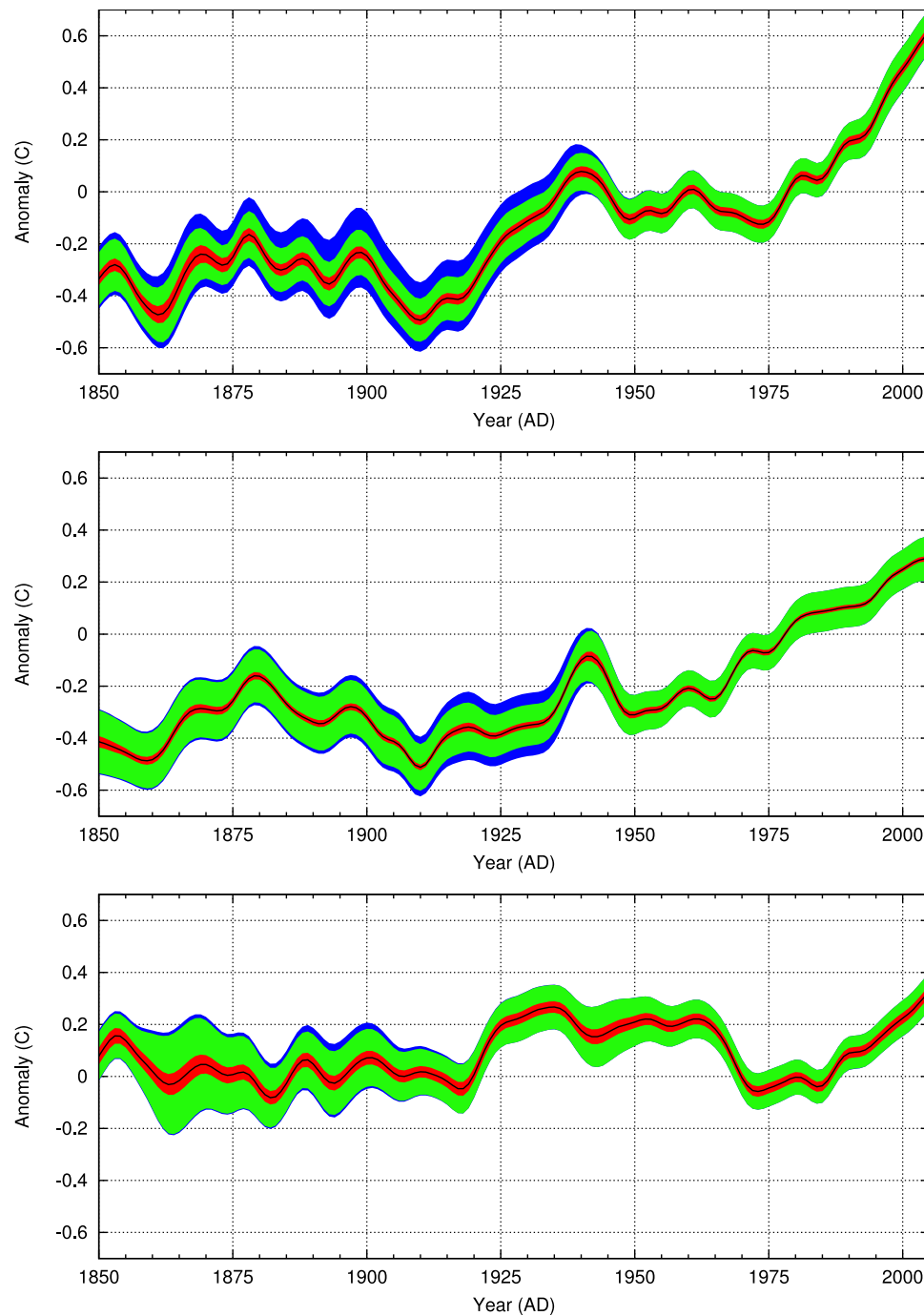


Figure 11. HadCRUT3 hemisphere temperature anomaly time series ($^{\circ}\text{C}$): (top) Northern Hemisphere (NH), (middle) Southern Hemisphere (SH), and (bottom) difference (NH – SH). The black line is the best estimate value; the red band gives the 95% uncertainty range caused by station, sampling, and measurement errors; the green band adds the 95% error range due to limited coverage; and the blue band adds the 95% error range due to bias errors.

[75] There are much larger uncertainties in the land data because the surface air temperature over land is much more variable than the SST. SSTs change slowly and are highly correlated in space; but the land air temperature at a given station has a lower correlation with regional and global temperatures than a point SST measurement, because land air temperature (LAT) anomalies can change rapidly in both time and space. This

means that one SST measurement is more informative about large-scale temperature averages than one LAT measurement. This difference also shows in the hemispheric differences (Figure 11): The Southern Hemisphere (SH) series has a similar uncertainty to the Northern Hemisphere (NH) series despite there being many more observations in the NH. This is because a larger fraction of the SH is sea, so fewer observations are needed.

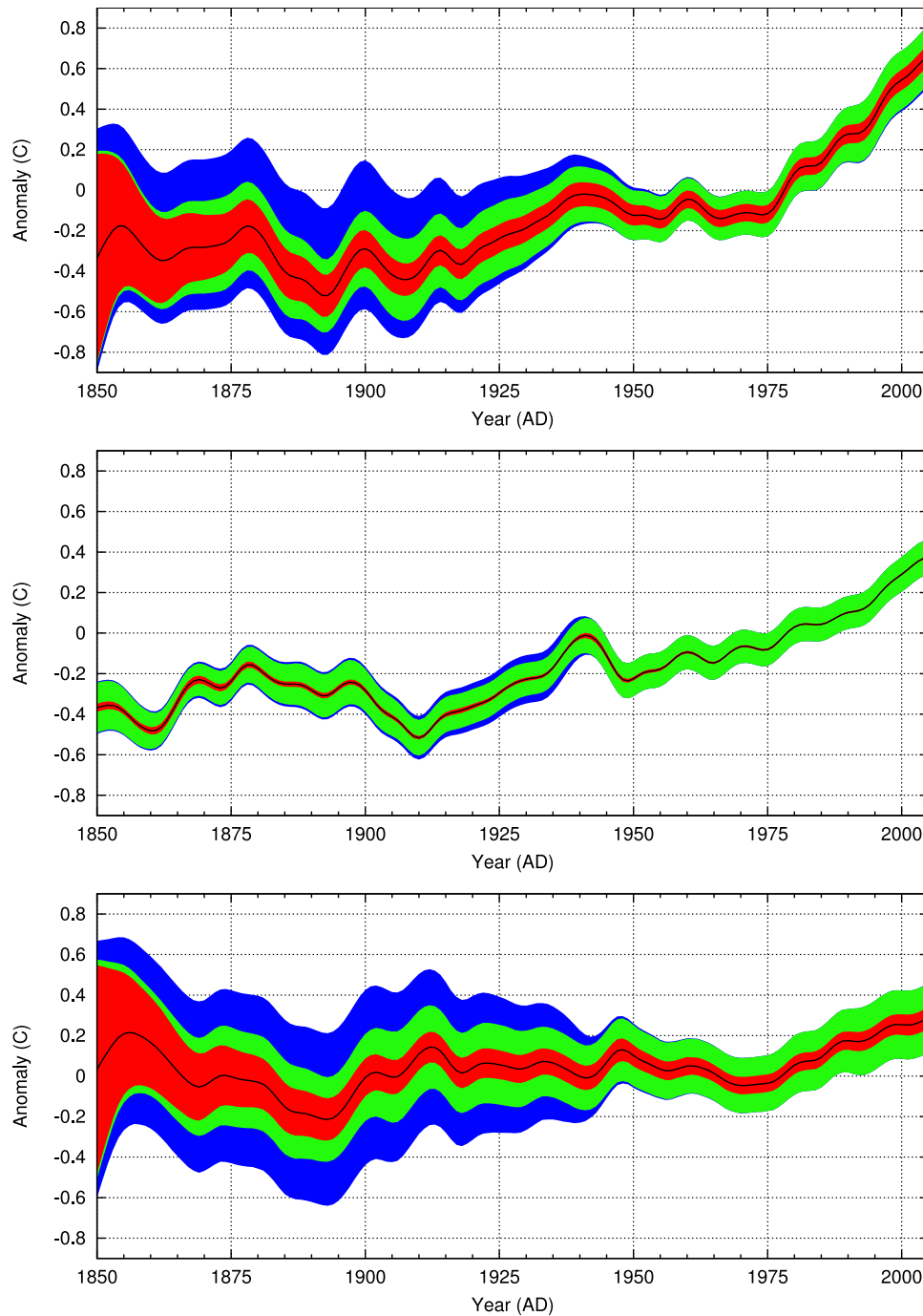


Figure 12. Global average of land and marine components of HadCRUT3 ($^{\circ}\text{C}$): (top) land, (middle) sea, and (bottom) difference (land – sea). The black line is the best estimate value; the red band gives the 95% uncertainty range caused by station, sampling, and measurement errors; the green band adds the 95% error range due to limited coverage; and the blue band adds the 95% error range due to bias errors.

[76] The difference between the land and sea temperatures (Figure 12, bottom) is not distinguishable from zero until about 1980. There are several possible causes for the recent increase: It could be a real effect, the land warming faster than the ocean (this is an expected response to increasing greenhouse gas concentrations in the atmosphere [Barnett *et al.*, 2000], but it could also indicate a change in the atmospheric circulation [Parker *et al.*, 1994]), it could indicate an uncorrected bias in one

or both data sources [see Rayner *et al.*, 2006, section 6], or it could be a combination of these effects. These issues have not been pursued further here, but such studies will form part of future work on land and marine temperatures and their uncertainties.

6.3. Comparison of Global Time Series With Previous Versions

[77] Figure 13 shows time series of the global average of the land data, the marine data, and the blended data set with

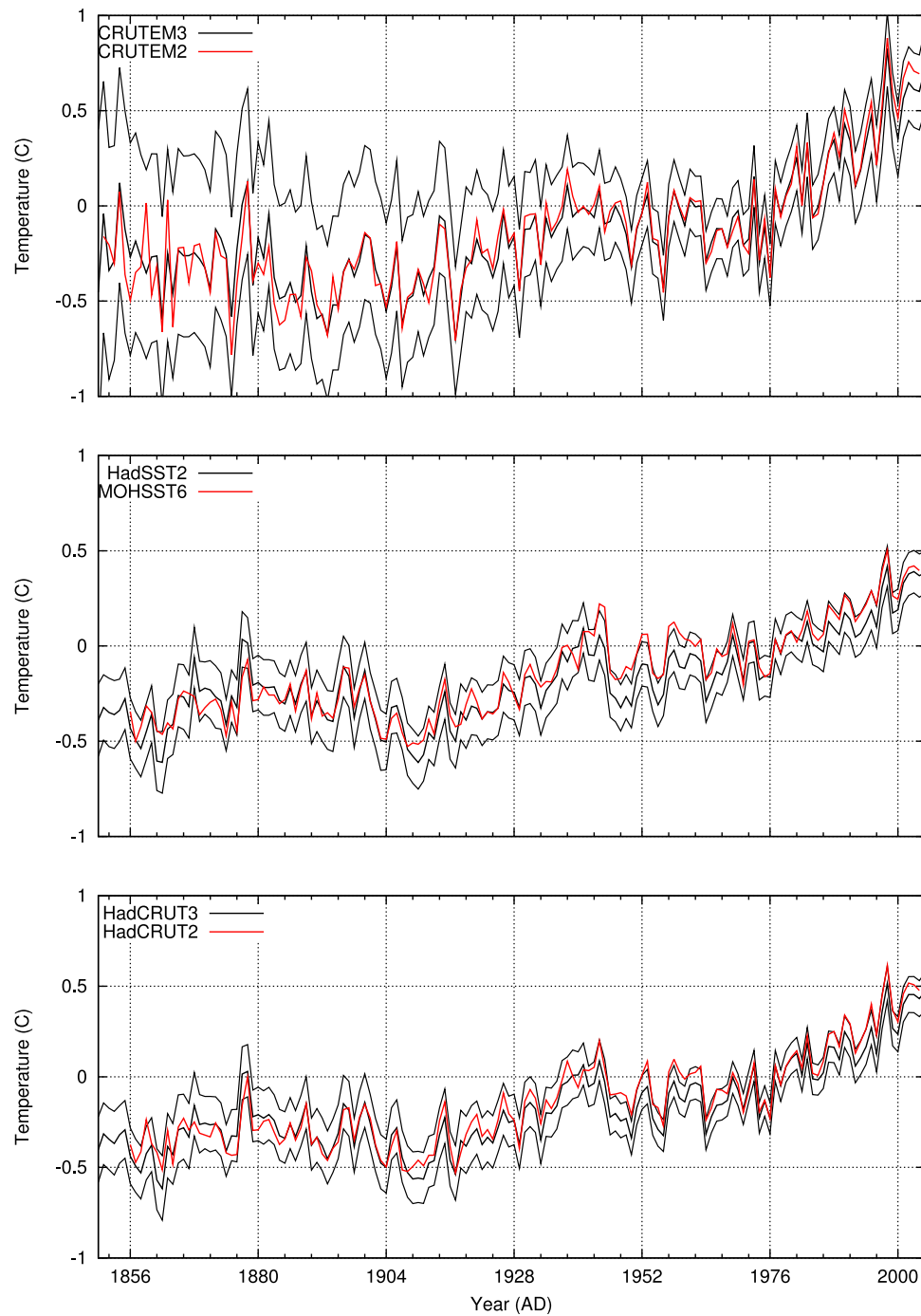


Figure 13. New data set versions and their 95% uncertainty ranges (in blue), compared with the previous version of each data set (in red): (top) land data, (middle) marine data, and (bottom) combined data.

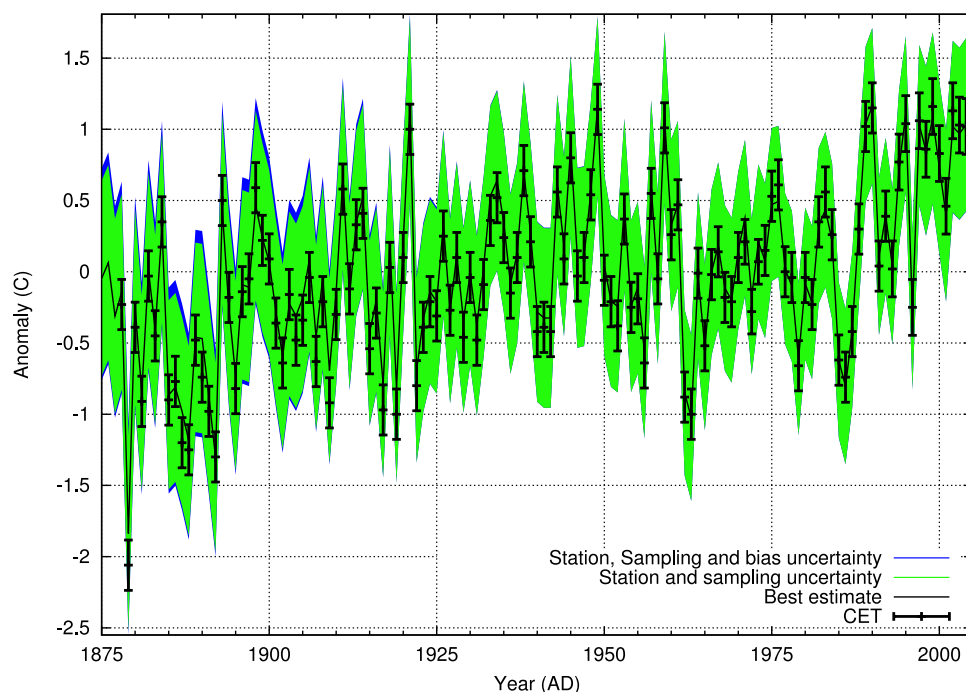


Figure 14. CRUTEM3 (for 50–55°N, 0–5°W) comparison with Central England Temperature (CET) (error ranges are 95%).

their uncertainty ranges, and compares them to the previous versions of each data set.

[78] The additions and improvements made to the land data do not make any large differences to the global land average, except very early in the record where the uncertainties are large. The new marine data, however, do produce some sizable changes: Refinements to the climatology have produced an offset, and new data have produced some other secular changes in the series.

[79] The differences between the old and new marine data series are sometimes outside the error range of the new series. Most of the difference is a constant offset due to changes to the climatology, and uncertainties in the climatology are not part of the error model for the marine data. (In the land data climatologies are estimated for each station, and as the mix of stations in any one grid box changes with time so does the climatology. So uncertainties in the station climatology are a component of the uncertainty in changes of gridded land temperature anomalies. However, for the marine data, climatologies are specified for each grid box, and they are constant in time, so uncertainties in the marine climatology do not contribute directly to uncertainties in changes in marine temperature anomalies). However, even after removing the constant offset produced by the climatology change, there are still differences between the old and new SST series that are larger than the assessed random and sampling errors. These differences suggest the presence of additional error components in the marine data. At the moment, the nature of these error components is not known for certain, but the main difference between the old and new data sets is the use of different sets of observations [Rayner *et al.*, 2006]. It seems likely that

different groups of observations may be measuring SST in different ways even in recent decades, and therefore there may be unresolved bias uncertainties in the modern data. Quantifying such effects will be a priority in future work on marine data.

6.4. Comparison With Central England Temperature

[80] The Central England Temperature (CET) series is the longest instrumental temperature record in the world [Parker *et al.*, 1992]. It records the temperature of a triangular portion of England bounded by London, Herefordshire and Lancashire, and provides mean daily temperature estimates back to 1772. The HadCRUT3 and CET series do use some of the same stations, but of the 13 sites that make some contribution to CRUTEM3 in the CET region, no more than 2 also contribute to CET, and there are always also stations contributing to CET but not to CRUTEM3. So if CET corresponds closely to the HadCRUT3 value for the central England grid box, it suggests that both series are correctly describing the local temperature changes, and is not simply a consequence of shared inputs. Recently, uncertainty estimates have been derived for CET since 1878 [Parker and Horton, 2005].

[81] The area covered by CET is less than 1 grid box in the $5^\circ \times 5^\circ$ gridded CRUTEM3 data set. Comparing the CET data with the corresponding grid box in CRUTEM3 (Figure 14) shows encouraging agreement: Despite being based on largely different observations, the two series agree within their uncertainties.

[82] Doing the same comparison using the full HadCRUT3 data (blended land and sea) gives a different picture (Figure 15). The $5^\circ \times 5^\circ$ grid box covering the CET region also contains much of the Irish Sea and the

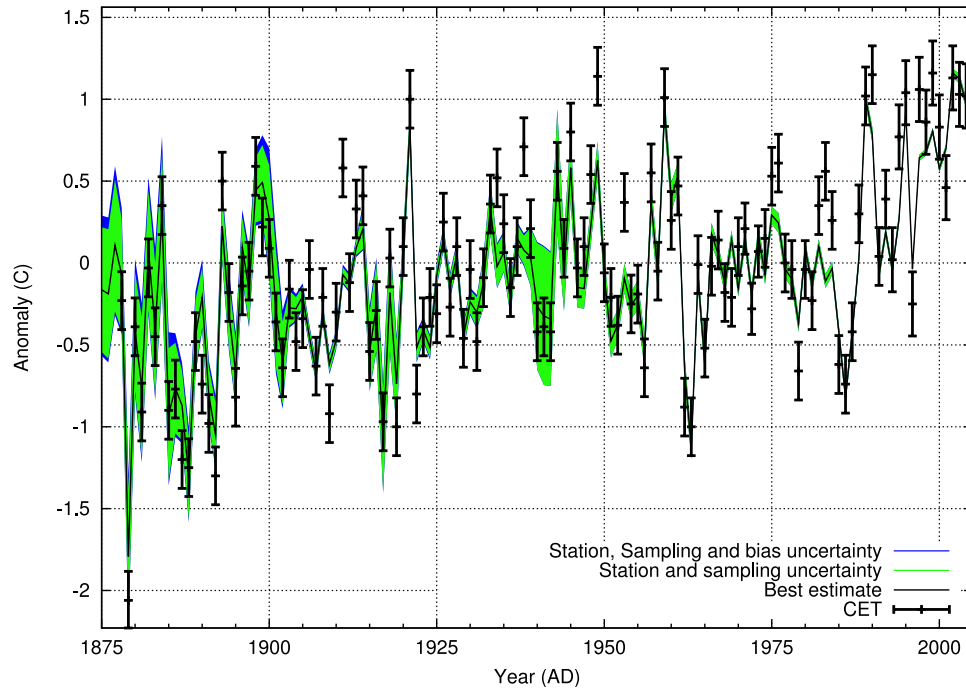


Figure 15. HadCRUT3 (for 50–55°N, 0–5°W) comparison with CET (error ranges are 95%).

English Channel; both regions where there are many SST observations. Many SST observations mean that the uncertainty on the SST monthly means is small, so the blended value is biased toward SST and has a small uncertainty.

[83] Adding the SST data has reduced the agreement with CET; and the uncertainty in the HadCRUT3 value is much smaller than the CRUTEM3 uncertainty because there are a lot of SST observations around the British coast. The uncertainty varies in time because, unlike the land data, the number of SST observations changes with time: The uncertainty increases in the early part of the series and during the two world wars are quite noticeable. Figure 15 demonstrates that the land and sea temperature anomalies in one $5^\circ \times 5^\circ$ grid box can have sizable differences in their annual values, although the longer-term changes are very similar.

[84] Because of these land-sea differences it will sometimes be better to use the land and sea specific data rather than the blended data set. For example when looking at paleodata from tree rings near coasts it is probably better to use the land data set CRUTEM3 than the blended data set HadCRUT3. Similarly for paleodata from coastal corals the SST data set should be used.

7. Conclusions

[85] A new version of the gridded historical surface temperature data set HadCRUT3 has been produced. This data set is a collaborative product of scientists at the Met Office Hadley Centre (who provide the marine data), and at the Climatic Research Unit at the University of East Anglia (who provide the land surface data). The new data set benefits from the improvements to the marine data de-

scribed by Rayner *et al.* [2006] as well as the improvements to the land data described in this paper. However, the principal advance over previous versions of the data set [Jones *et al.*, 2001; Jones and Moberg, 2003] is in the provision of a comprehensive set of uncertainties to accompany the gridded temperature anomalies.

[86] As well as variance adjustments (adjustments to the data to allow for the changing numbers of observations), fields of measurement and sampling, and bias uncertainty have been produced. All the gridded data sets, and some time series derived from them, are available from the Web sites <http://www.hadobs.org> and <http://www.cru.uea.ac.uk>.

[87] The gridded data sets start in 1850 because there are too few observations available from before this date to make a useful gridded field. Many marine observations from the first half of the nineteenth century are known to exist in log books kept in the British Museum and the U.K. National Archive, but these observations have never been digitized. If these observations were available, it is likely that the gridded data sets, and so information on surface climate change and variability, could be extended by several decades.

Appendix A: Variance Adjustment Method

A1. Description of the Method

[88] The relationship between the variance in a grid box and the variance of individual station observations is given by Jones *et al.* [1997]

$$\sigma_n^2 = \frac{\bar{\sigma}_i^2(1 + (n-1)\bar{r})}{n}, \quad (\text{A1})$$

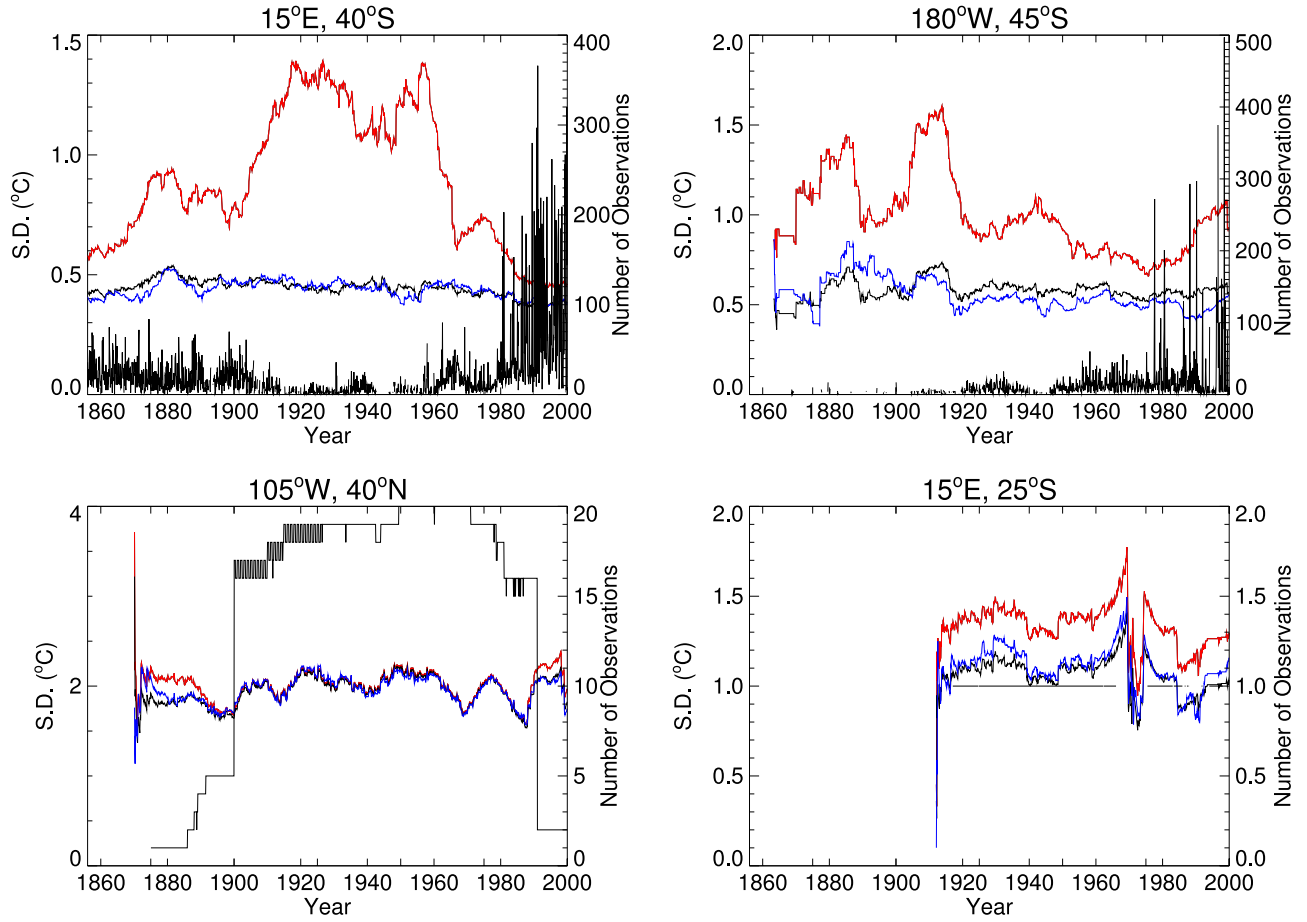


Figure A1. Running 10-year standard deviations ($^{\circ}\text{C}$, left axis scale) are shown for four grid boxes. The blue line shows the standard deviation of the perfect model data masked to have the same coverage as the data. The red line shows the standard deviation of the synthetic data before variance adjustment, and the black line shows the standard deviation of the synthetic data after variance adjustment. The number of observations is also shown (right-hand scale). The top two panels are marine grid boxes, the lower two are land grid boxes.

where σ_n^2 is the variance of the grid box average, $\bar{\sigma}_i^2$ is the mean variance of the individual station time series that contribute to that grid box average and \bar{r} is the average correlation of stations within the grid box. Two interesting variables can be derived from this. The first is the true grid box variance, $\sigma_{n=\infty}^2$. That is the variance the grid box average would have if it contained an infinite number of observations

$$\sigma_{n=\infty}^2 = \bar{\sigma}_i^2 \bar{r}. \quad (\text{A2})$$

The second is the sampling error, m_s^2 , equal to the difference between equations (A1) and (A2)

$$m_s^2 = \frac{\bar{\sigma}_i^2(1 - \bar{r})}{n}. \quad (\text{A3})$$

Equation (A1) assumes that the time series of the grid box anomaly is stationary. In fact, the average temperature in an area defined by a grid box exhibits natural variability on a variety of timescales: a long-term trend

(perhaps because of global warming), interdecadal variability (perhaps because of modes like ENSO) and higher-frequency natural variability. To ensure that the series is stationary, the anomalies in individual grid boxes were detrended using a 6-year running average centered on the month of interest.

[89] The detrended anomalies were then multiplied by an adjustment factor,

$$k = \sqrt{\frac{\sigma_{n=\infty}^2}{m_{total}^2(n, t) + \sigma_{n=\infty}^2}}, \quad (\text{A4})$$

where m_{total}^2 is the estimated random error, a combination of sampling, measurement and other errors, expressed as a function of the number of observations, n , and time, t . For marine data m_{total}^2 and $\sigma_{n=\infty}^2$ were as calculated by Rayner *et al.* [2006]. For land data, values for m_{total}^2 were calculated as in section 2.3.2 and the values of $\sigma_{n=\infty}^2$ were calculated from equation (A2) using the individual station variances and the average correlations between them.

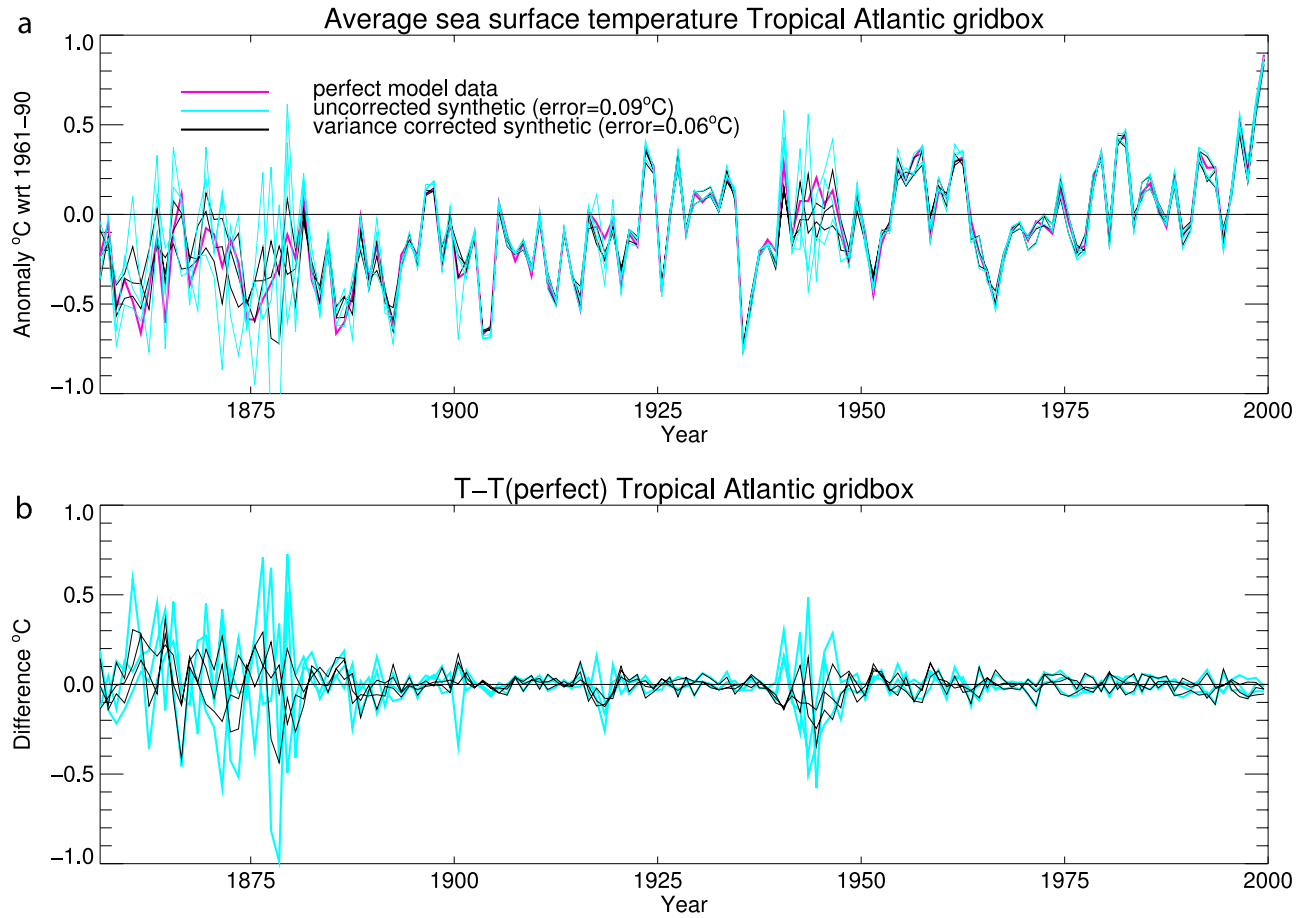


Figure A2. (a) Annual average sea surface temperatures from a grid box in the tropical Atlantic for the original model data (magenta) and three realizations of the synthetic data before (cyan) and after (black) variance adjustment. (b) Difference between the unadjusted synthetic data and model data (cyan) and the difference between the variance adjusted synthetic data and model data (black).

After the adjustment factor was applied, the smoothed series was added back to recover the variance adjusted time series.

A2. Test of the Method

[90] If it is working well, variance adjustment should reduce the random noise in the temperature values introduced by having only a limited number of observations, but leave the real underlying temperature variations unchanged. This cannot be tested using the actual HadCRUT3 data, as the distinction between real variations and noise is unknown. To test the method, the pseudo-proxy method of *von Storch et al.* [2004] has been adapted to instrumental data. A synthetic version of HadCRUT3 has been made by adding noise to sub-sampled GCM temperature data; the test is then to see how well variance adjustment recovers the original GCM data from the synthetic HadCRUT3.

A2.1. Making the Synthetic Data Set

[91] A synthetic data set was constructed using an all forcings run [Tett et al., 2006] of the HadCM3 [Gordon et al., 2000] GCM. Values of \bar{r} for land data grid boxes were calculated for the detrended model data following

the procedure of *Jones et al.* [1997]. In marine grid boxes, values of \bar{r} were calculated in both time and space to take into account the fact that marine observations are point measurements rather than monthly averages as in the land data. The time component was calculated by fitting an exponential to the lagged correlations of monthly anomalies in a given grid box and using the fitted correlation decay time to estimate the average correlation across the grid box. These were used to calculate estimated station variances by assuming that the variance of the model temperature anomalies in a grid box represented the variance in that grid box for an infinite number of stations, $\sigma_{n=\infty}^2$. In this instance the value of σ_i^2 can be easily extracted from equation A2. These average station variances were then used to create a synthetic time series for each grid box that showed variance fluctuations of a kind seen in the observational data. The variance of the time series was inflated by adding random noise of variance, v^2 , calculated using

$$v^2 = \frac{\bar{\sigma}_i^2(1 - \bar{r}) + m_m^2}{n}, \quad (\text{A5})$$

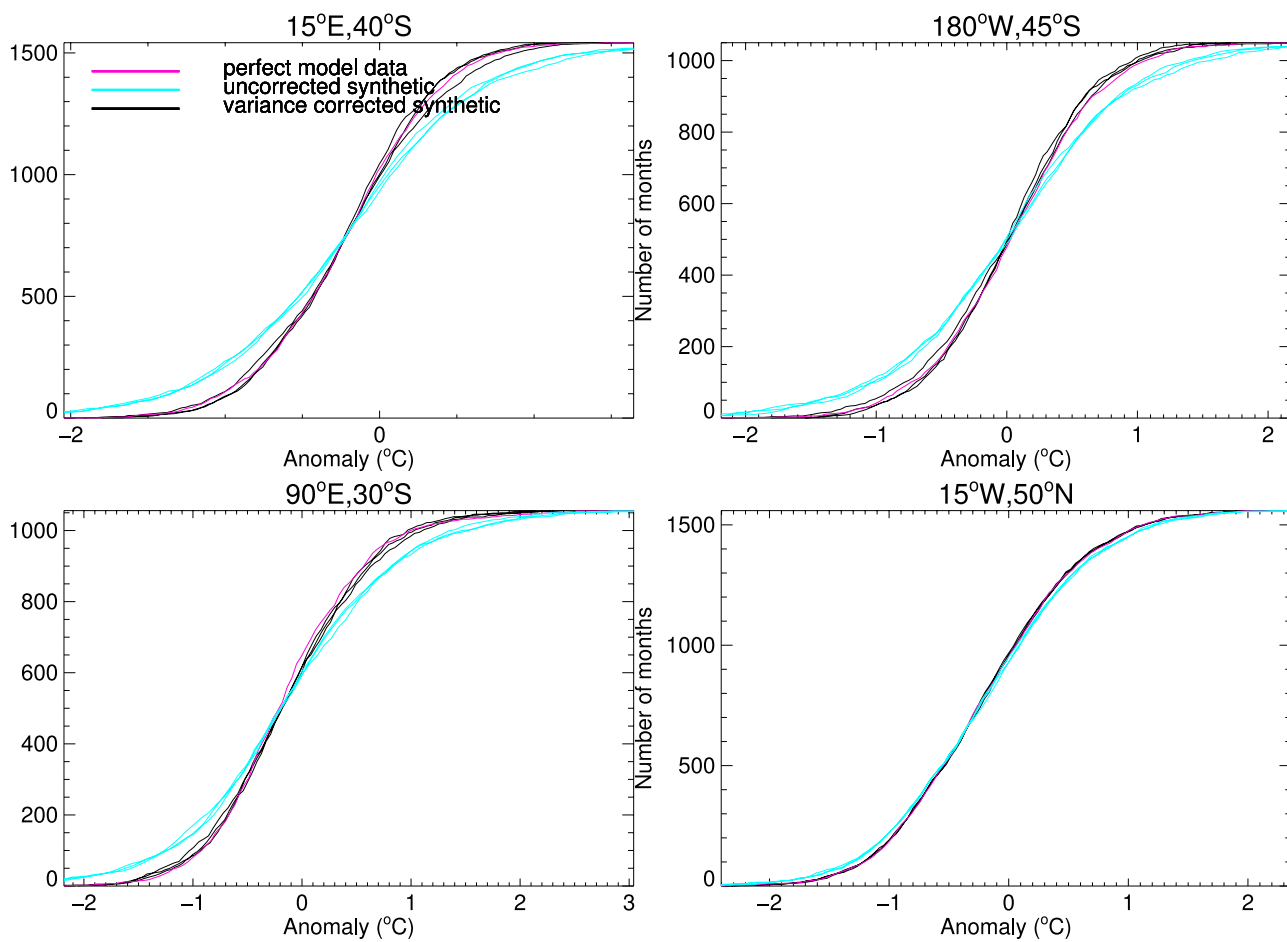


Figure A3. Cumulative frequency distributions of monthly anomalies in four marine grid boxes. The magenta lines show the original model data, the cyan lines show the three realizations of the unadjusted synthetic data, and the black lines show the three realizations of the variance-adjusted synthetic data.

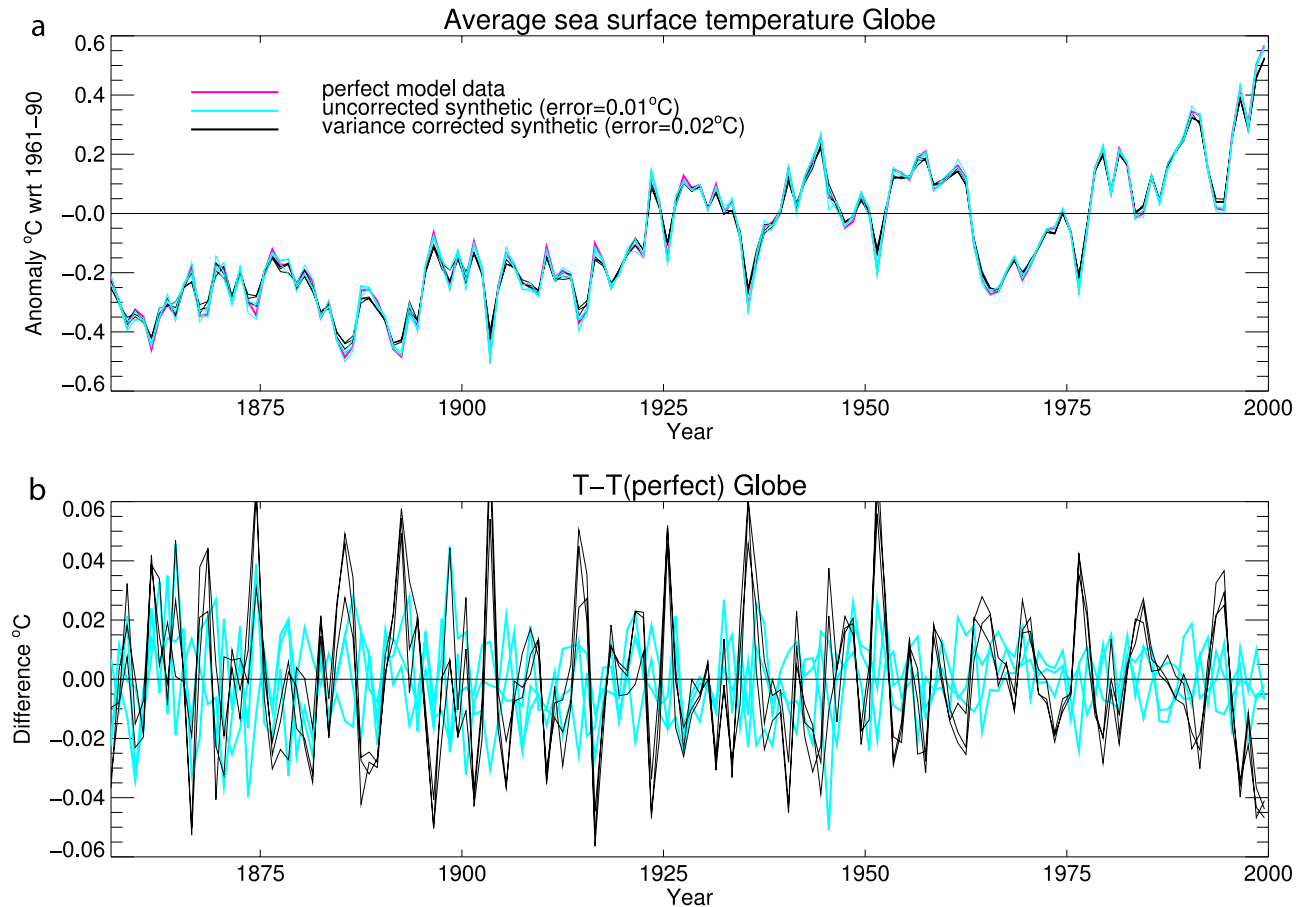


Figure A4. (a) Annual average sea surface temperatures from the whole globe for the original model data (magenta) and three realizations of the synthetic data before (cyan) and after (black) variance adjustment. (b) Difference between the unadjusted synthetic data and model data (cyan) and the difference between the variance-adjusted synthetic data and model data (black).

where the n were a realistic distribution of numbers of observations as obtained from the historical records of monthly average temperatures; m_m^2 was an estimate of the measurement error, which was assumed to be negligible over land. Three realizations of the synthetic data were created. They differed only in the random numbers used to generate the random noise which was added to the time series.

A2.2. Comparing Adjusted and True Data

[92] The synthetic data were then run through the variance adjustment algorithms and the variance of the output was compared to that of the original model data (see Figure A1). Before variance adjustment the variance of an average land data grid box was overestimated by around 11% and the variance of an average marine grid box by 180%. After variance adjustment the variance of an average land data grid box was found to be underestimated by less than 2% and the variance of an average marine data grid box was underestimated by 5%. In the marine case, discrepancies from the true variance can be larger than this in individual grid boxes, although in all cases the adjusted variance is closer to the true value than the unadjusted variance.

[93] In individual grid boxes variance adjustment typically brings the synthetic data closer to the true value (see, for example, Figure A2), especially at times when such adjustments are large. This is notable, for example, during the second world war or early in the record. The frequencies of individual grid box monthly averages are also typically improved (see, for example, Figure A3) with extreme outliers because of noise being effectively adjusted. This means that it is possible to make more meaningful analyses of the occurrences of true extremes using the variance adjusted data.

[94] However, when these individual variance-adjusted grid box values are averaged over large regions (Figure A4), the opposite is true. Whereas the random errors of individual grid boxes tend to cancel out when averaged, the cumulative effect of the hundreds of slight, but correlated, variance adjustments is to reduce the variance of the regional average.

[95] Some degradation of the true temperature signal is inevitable, as no filter can perfectly separate out the measurement and sampling error component of the temperature signal, and the reduction applied to the noise component will then be applied to some of the signal as well. Despite

this, the variance adjustment process is very successful at the grid box scale.

[96] **Acknowledgments.** This work was funded by the Public Met. Service Research and Development Contract and by the Department for Environment, Food and Rural Affairs, under contract PECD/7/12/37. The development of the basic land station data set has been supported over the last 27 years by the Office of Science (BER), U.S. Department of Energy, most recently by grant DE-FG02-98ER62601. The authors are grateful to Tom Peterson, of the U.S. National Climatic Data Centre, for many valuable suggestions.

References

- Barnett, T. P., G. Hegerl, T. Knutson, and S. Tett (2000), Uncertainty levels in predicted patterns of anthropogenic climate change, *J. Geophys. Res.*, **105**, 15,525–15,542.
- Begert, M., T. Schlegel, and W. Kirchhofer (2005), Homogeneous temperature and precipitation series of Switzerland for 1864 to 2000, *Int. J. Climatol.*, **25**, 65–80.
- Diaz, H., C. Folland, T. Manabe, D. Parker, R. Reynolds, and S. Woodruff (2002), Workshop on advances in the use of historical marine climate data, *WMO Bull.*, **51**, 377–380.
- Folland, C. K., and D. E. Parker (1995), Correction of instrumental biases in historical sea surface temperature data, *Q. J. R. Meteorol. Soc.*, **121**, 319–367.
- Folland, C. K., D. E. Parker, and T. N. Palmer (1986), Sahel rainfall and worldwide sea temperatures 1901–85, *Nature*, **320**, 602–607.
- Folland, C. K., et al. (2001), Global temperature change and its uncertainties since 1861, *Geophys. Res. Lett.*, **28**(13), 2621–2624.
- Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood (2000), The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments, *Clim. Dyn.*, **16**, 147–168.
- Intergovernmental Panel on Climate Change (2001), *Climate Change 2001: The Scientific Basis: Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, edited by J. T. Houghton et al., 881 pp., Cambridge Univ. Press, New York.
- Johns, T., et al. (2004), HadGEM1—Model description and analysis of preliminary experiments for the IPCC Fourth Assessment Report, *Tech. Note 55*, Hadley Cent., Exeter, U.K.
- Jones, P. D. (1994), Hemispheric surface air temperature variations: A reanalysis and an update to 1993, *J. Clim.*, **7**, 1794–1802.
- Jones, P. D., and A. Moberg (2003), Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001, *J. Clim.*, **16**, 206–223.
- Jones, P. D., S. C. B. Raper, B. D. Santer, B. S. G. Cherry, C. M. Goodess, P. M. Kelly, T. M. L. Wigley, R. S. Bradley, and H. F. Diaz (1985), A grid point surface air temperature data set for the Northern Hemisphere, *Tech. Rep. TR022*, 251 pp., Carbon Dioxide Res. Div., Off. of Basic Energy Serv., U.S. Dep. of Energy, Washington, D. C.
- Jones, P. D., S. C. B. Raper, B. S. G. Cherry, C. M. Goodess, and T. M. L. Wigley (1986) A grid point surface air temperature data set for the Southern Hemisphere, 1851–1984, *Tech. Rep. TR027*, 73 pp., Carbon Dioxide Res. Div., U.S. Dep. of Energy.
- Jones, P. D., P. Y. Groisman, M. Coughlan, N. Plummer, W. C. Wang, and T. R. Karl (1990), Assessment of urbanization effects in time series of surface air temperature over land, *Nature*, **347**, 169–172.
- Jones, P. D., T. J. Osborn, and K. R. Briffa (1997), Estimating sampling errors in large-scale temperature averages, *J. Clim.*, **10**, 2548–2568.
- Jones, P. D., T. J. Osborn, K. R. Briffa, C. K. Folland, E. B. Horton, L. V. Alexander, D. E. Parker, and N. A. Rayner (2001), Adjusting for sampling density in grid box land and ocean surface temperature time series, *J. Geophys. Res.*, **106**, 3371–3380.
- Kalnay, E., and M. Cai (2003), Impact of urbanization and land-use change on climate, *Nature*, **423**, 528–531.
- Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, **77**, 437–471.
- Kerr, R. A. (2005), Atlantic climate pacemaker for millennia past, decades hence?, *Science*, **309**, 5731, 41–43.
- Manabe, T. (2003), The Kobe collection: Newly digitized Japanese historical surface marine meteorological observations, in *Advances in the Applications of Marine Climatology—The Dynamic Part of the WMO Guide to the Applications of Marine Meteorology*, WMO/TD 1081, 246 pp., World Meteorol. Organ., Geneva, Switzerland.
- Parker, D. E. (1994), Effects of changing exposure of thermometers at land stations, *Int. J. Climatol.*, **14**, 1–31.
- Parker, D. E. (2004), Large-scale warming is not urban, *Nature*, **432**, 290.
- Parker, D. E., and B. Horton (2005), Uncertainties in Central England Temperature 1873–2003 and some updates to the maximum and minimum series, *Int. J. Climatol.*, **25**, 1173–1188.
- Parker, D. E., T. P. Legg, and C. K. Folland (1992), A new daily Central England Temperature series, *Int. J. Climatol.*, **12**, 317–342.
- Parker, D. E., P. D. Jones, A. Bevan, and C. K. Folland (1994), Interdecadal changes of surface temperature since the 19th century, *J. Geophys. Res.*, **99**, 14,373–14,399.
- Parker, D. E., C. K. Folland, and M. Jackson (1995), Marine surface temperature: Observed variations and data requirements, *Clim. Change*, **31**, 559–600.
- Peterson, T. C. (2004), Assessment of urban versus rural in situ surface temperatures in the contiguous United States: No difference found, *J. Clim.*, **16**, 2941–2959.
- Peterson, T. C., and T. W. Owen (2005), Urban heat island assessment: Metadata are important, *J. Clim.*, **18**, 2637–2646.
- Rayner, N. A., P. Brohan, D. E. Parker, C. K. Folland, J. J. Kennedy, M. Vanicek, T. J. Ansell, and S. F. B. Tett (2006), Improved analyses of changes and uncertainties in sea-surface temperature measured in-situ since the mid-nineteenth century, *J. Clim.*, **19**, 446–469.
- Simmons, A. J., P. D. Jones, V. da Costa Bechtold, A. C. M. Beljaars, P. W. Källberg, S. Saarinen, S. M. Uppala, P. Viterbo, and N. Wedi (2004), Comparison of trends and low-frequency variability in CRU, ERA-40 and NCEP/NCAR analyses of surface air temperature, *J. Geophys. Res.*, **109**, D24115, doi:10.1029/2004JD005306.
- Tett, S. F. B., R. Betts, T. J. Crowley, T. Johns, A. Jones, J. Gregory, T. Osborn, E. Öström, D. L. Roberts, and M. J. Woodage (2006), The impact of natural and anthropogenic forcings on climate since 1550, *Clim. Dyn.*, in press.
- Turner, J., S. R. Colwell, G. J. Marshall, T. A. Lachlan-Cope, A. M. Carleton, P. D. Jones, V. Lagun, P. A. Reid, and S. Iagovkina (2005), Antarctic climate change during the last 50 Years, *Int. J. Climatol.*, **25**, 279–294.
- Vincent, L. A., and D. W. Gullet (1999), Canadian historical and homogeneous temperature datasets for climate change analyses, *Int. J. Climatol.*, **19**, 1375–1388.
- von Storch, H., E. Zorita, J. Jones, Y. Dimitriev, F. González-Rouco, and S. Tett (2004), Reconstructing past climate from noisy data, *Science*, **306**, 679–682.
- World Meteorological Organization (1996), *Climatological normals (CLINO) for the period 1961–1990*, WMO/OMM 847, 768 pp., Geneva, Switzerland.
- Woodruff, S. D., S. J. Worley, J. A. Arnott, H. F. Diaz, J. D. Elms, M. Jackson, S. J. Lubker, and D. E. Parker (2003), COADS updates and the blend with the U.K. Met Office Marine Data Bank, in *Advances in the Applications of Marine Climatology—The Dynamic Part of the WMO Guide to the Applications of Marine Meteorology*, WMO/TD 1081, 246 pp., World Meteorol. Organ., Geneva, Switzerland.
- Zhou, L., R. E. Dickinson, Y. Tian, J. Fang, Q. Li, R. K. Kaufmann, C. J. Tucker, and R. B. Myneni, (2004), Evidence for a significant urbanization effect on climate in China, *PNAS*, **101**, 26, 9540–9544.

P. Brohan and J. J. Kennedy, Hadley Centre for Climate Prediction and Research, Met Office, Exeter EX1 3PB, UK. (philip.brohan@metoffice.gov.uk)

I. Harris and P. D. Jones, Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

S. F. B. Tett, Met Office Hadley Centre (Reading Unit) Meteorology Building, University of Reading, Reading RG6 6BB, UK.